

CRFE

CRFE

Access DB# 84868

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: David Fox Examiner #: 65401 Date: 1/21/03
Art Unit: 1638 Phone Number 308-0280 Serial Number: 09/701,023
Mail Box and Bldg/Room Location: CM1 Results Format Preferred (circle): PAPER DISK E-MAIL
9E12 9E15 9E15

If more than one search is submitted, please prioritize searches in order of need. mej

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc. if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____ Point of Contact: Susan Hanley
Inventors (please provide full names): _____ Technical Info. Specialist CM1 6B05 Tel: 305-4053

Earliest Priority Filing Date: _____ renewed 2/2/03

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please do sequence search for
SEQ ID NO: 4 (protein) and
SEQ ID NO: 1, nucleotides 81-1024
gene encoding it). best 3/3/9

Amendment due 2/8/03

1A
1-1302
4-~~1302~~ 314
THANK YOU

David

TAFF USE ONLY

Type of Search		Vendors and cost where applicable
Searcher: <u>Hanley</u>	NA Sequence (#) <u>1</u>	STN _____
Searcher Phone #: _____	AA Sequence (#) <u>1</u>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>1/22</u>	Bibliographic _____	Dr.Link _____
Date Completed: <u>1/27</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: <u>2</u>	Fulltext _____	Sequence Systems <u>01</u>
Technical Prep Time: <u>3</u>	Patent Family _____	WWW/Internet _____
Online Time: <u>7</u>	Other _____	Other (specify) _____

This Page Blank (uspto)

M nucleic - nucleic search, using sw model

un on: January 25, 2003, 21:54:27 ; Search time 2843 seconds
(without alignments)
9663.401 Million cell updates/sec

file: US-09-701-023-1_COPY_81_1024
effect score: 944
sequence: 1 tggcgactctctctcttc.....attgtctctgaagcttttaa 944

coring table: IDENTRY_NUC

Gapop 10.0 , Gapext 1.0

searched: 2054640 seqs, 14551402878 residues

total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.ov.*

5: gb.pat.*

6: gb.ph.*

7: gb.pl.*

8: gb.pr.*

9: gb.ro.*

10: gb.sit.*

11: gb.sy.*

12: gb.un.*

13: gb.vi.*

14: em.ba.*

15: em.fun.*

16: em.in.*

17: em.mu.*

18: em.or.*

19: em.ov.*

20: em.pat.*

21: em.ph.*

22: em.pl.*

23: em.ro.*

24: em.sit.*

25: em.sy.*

26: em.un.*

27: em.vi.*

28: em.htg_hum.*

29: em.htg_inv.*

30: em.htg_other.*

31: em.htg_mus.*

32: em.htg_pln.*

33: em.htg_rod.*

34: em.htg_vrt.*

35: em.htg_mam.*

36: em.htg_vrt.*

37: em.htg_mus.*

38: em.htg_hum.*

39: em.htgo_mus.*

40: em.htgo_hum.*

41: em.htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	944	100.0	1302	8	AF159255	Arabidops
2	736	78.0	1133	8	AF146794	Arabidops
3	732.8	77.6	82381	8	ATF27619	Arabidops
4	732.8	77.6	83371	8	ATM4122	Arabidops
5	732.8	77.6	198372	8	ATCHRIV66	Arabidops
6	45.8	4.9	7218	6	I66494	Sequence 14
7	41.8	4.4	141632	2	AC104900	Mus muscu
8	41	4.3	119072	9	ALI36531	Human DNA
9	41	4.3	203466	2	AC108536	Rattus no
10	40.6	4.3	158717	9	ALI33412	Human DNA
11	40.4	4.3	1472	1	SEMP	Streptococ
12	40.4	4.3	4826	1	SDY18363	Streptococ
13	40.4	4.3	116688	9	HSJ393E18	Human DNA
14	40.2	4.3	153168	9	AC011974	Homo sapi
15	40.2	4.3	180895	2	AC110552	Mus muscu
16	40	4.2	58221	9	AC093769	Homo sapi
17	39.8	4.2	131903	9	HS360E18	Human DNA
18	39.8	4.2	139843	2	ALI58145	Homo sapi
19	39.6	4.2	70040	2	AC124983	Mus muscu
20	39.6	4.2	97386	9	AC037453	Homo sapi
21	39.6	4.2	143125	9	AC092338	Homo sapi
22	39.6	4.2	178229	2	AL713999	Homo sapi
23	39.2	4.2	81296	2	AC094193	Rattus no
24	39.2	4.2	286175	2	AC094286	Rattus no
25	39	4.1	550	11	HUMUT5350	Human chrom
26	39	4.1	162921	2	AC093367	Homo sapi
27	39	4.1	194407	2	AC068845	Homo sapi
28	38.8	4.1	10503	1	AE010877	Methanosa
29	38.8	4.1	77335	2	PFMAL13P2_3	Continuation (4 of
30	38.8	4.1	218369	2	AC122351	Mus muscu
31	38.6	4.1	103875	9	ALI36223	Human DNA
32	38.6	4.1	147598	9	AC093503	Homo sapi
33	38.4	4.1	93187	2	AC092769	Homo sapi
34	38.4	4.1	152711	9	AP001178	Homo sapi
35	38.4	4.1	166816	2	AC127328	Mus muscu
36	38.4	4.1	188682	2	AC102874	Mus muscu
37	38.4	4.1	199342	2	AC110510	Mus muscu
38	38.4	4.1	205562	2	AL732497	Mus muscu
39	38.4	4.1	208083	2	AC009649	Homo sapi
40	38.4	4.1	248673	2	AC115361	Mus muscu
41	38.4	4.1	320679	2	AC126445	Mus muscu
42	38.2	4.0	194	11	GI2787	SWSS2144 Er
43	38	4.0	665	11	HUMUT7202	Human STS U
44	38	4.0	110819	9	AC055716	Homo sapi
45	38	4.0	148421	2	AC097306	Rattus no

ALIGNMENTS

RESULT 1
AF159255
LOCUS Arabidopsis thaliana 1302 bp mRNA linear PLN 08-SEP-1999
DEFINITION Arabidopsis thaliana sporocytelless (SPL) mRNA, complete cds.
ACCESSION AF159255
VERSION AF159255.1 GI:5566239
KEYWORDS
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 1302)
AUTHORS Yang, W.C., Ye, D., Xu, J. and Sundaresan, V.
TITLE The SPOROCYTELESS gene of Arabidopsis is required for initiation of

sporogenesis and encodes a novel nuclear protein
Genes Dev. 13 (16), 2108-2117 (1999)
MEDLINE 93396716
PUBMED 10463788
REFERENCE 2 (bases 1 to 1302)
AUTHORS Yang W.-C., Ye D., Xu J. and Sundaresan, V.
TITLE Direct Submission
SUBMITTED (14-JUN-1999) Plant Molecular Genetics, Institute of
Molecular Agrobiology, 1 Research Link, Singapore 117604, Singapore
FEATURES Location/Qualifiers

```

source
1..1302
/organism="Arabidopsis thaliana"
/cultivar="Landsberg erecta"
/db_xref="taxon:3702"
/chromosome="4"
/map="PG11-m123"
/note="nucleus"
1..1302
/gene="SPL"
80..1024
/gene="SPL"
/note="MADS-box related protein"
/codon_start=1
/product="sporocytelless"
/protein_id="AAD45344.1"
/db_xref="GI:5566240"
/translation="MATSLFFMTQDSVGNPNDLLRNLVNSSGEIRTEITLKSRG
RKPSGTGQOKKPTPLRGWVAKLERIEEKQLAAATVGDTSVAISNNATLAL
PVPDPCVVLQGPSSLGSRNRIYCGGVSGQVMDPVI SPGWETSTTHELSSNHL
POMENAVNNCRDTCFKKRLDGDNNVPSNGGFSKYTWIIPPMNGYDYLLOSDH
HORSQGLFDHRTARAASVSASSTTINPYNEATNHTGPMEEFGSYMEGNPRNGSGGV
KEYEFPFGYGERVSVVATSSVLVGDCSPNTIDLSQL"

BASE COUNT      402 a 273 c 275 g 352 t
ORIGIN

Query Match      100.0%; Score 944; DB 8; Length 1302;
Best Local Similarity 100.0%; Pred. No. 2.8e-260;
Matches 944; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Db	561	CTCAAAATGTTTAAAGCTTCTTCCAATAATACGCTGTGACACTTGCTTCAAGAGAAGACGTT	620
Qy	541	TGGATGTTGATCAGAAATAATGTTAGTTCGATCCAAACGGTGTGGATTTTCGAAATACACAA	600
Db	621	TGGATGTTGATCAGAAATAATGTTAGTTCGATCCAAACGGTGTGGATTTTCGAAATACACAA	680
Qy	601	TGATTCTCTCCGATCAAGCGGTACGATCAGTATCTTCTTCAATCAGATCATCATCAGA	660
Db	681	TGATTCTCTCCGATCAAGCGGTACGATCAGTATCTTCTTCAATCAGATCATCATCAGA	740
Qy	661	GGAGCCCAAGTTTCCTTTATGATCATAGAATCGCTAGAGCAGGTTTCAGTTTCTGCTTCTA	720
Db	741	GGAGCCCAAGTTTCCTTTATGATCATAGAATCGCTAGAGCAGGTTTCAGTTTCTGCTTCTA	800
Qy	721	GTACTACTATTAATCCTTATTTCAACGAGGCAACAATCATCGGACCACCAATGGAGGAAT	780
Db	801	GTACTACTATTAATCCTTATTTCAACGAGGCAACAATCATCGGACCACCAATGGAGGAAT	860
Qy	781	TTGGGAGCTACATGGAGGAAACCCCTAGAATGGATCAGAGGTGTGAAGGAGTACGAGT	840
Db	861	TTGGGAGCTACATGGAGGAAACCCCTAGAATGGATCAGAGGTGTGAAGGAGTACGAGT	920
Qy	841	TTTTTCCGGGGAATATGGTGAAGAGTTTTCAGTGTGGGTACAAGTCGTCACTCGTAG	900
Db	921	TTTTTCCGGGGAATATGGTGAAGAGTTTTCAGTGTGGGTACAAGTCGTCACTCGTAG	980
Qy	901	GTGATTGCAGTCCTAATACCAATTGATTTGCTTGAAGCTTTAA	944
Db	981	GTGATTGCAGTCCTAATACCAATTGATTTGCTTGAAGCTTTAA	1024

RESULT 2 .
AF146794 .
LOCUS AF146794 .
DEFINITION Arabidopsis thaliana NOZZLE (NZZ) gene, complete cds.
ACCESSION AF146794
VERSION AF146794.1 GI:5007008
KEYWORDS .
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 1133)
AUTHORS Schiefthaler,U., Balasubramanian,S., Sieber,P., Chevalier,D., Wisman,E. and Schneitz,K.
TITLE Molecular analysis of NOZZLE, a gene involved in pattern formation and early sporogenesis during sex organ development in Arabidopsis thaliana
JOURNAL Proc.Natl. Acad. Sci. U.S.A. 96 (20), 11664-11669 (1999)
MEDLINE 99432290
PUBMED 10500234
REFERENCE 2 (bases 1 to 1133)
AUTHORS Schiefthaler,U., Balasubramanian,S., Chevalier,D., Sieber,P. a Schneitz,K.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1999) Institute of Plant Biology, University Zurich, Zollikerstr. 107, Zurich, ZH 8008, Switzerland

```

FEATURES             Location/Qualifiers
     source            1..1133
                        /organism="Arabidopsis thaliana"
                        /cultivar="Landsberg erecta"
                        /db_xref="taxon:3702"
                        /chromosome="4"
     gene              /map="between PG11 and PRHA"
                        <1..>1133
     mRNA              /gene="NZZ"
                        join(<1..531,631..864,954..>1133)
                        /gene="NZZ"
     CDS               /product="NOZZLE"
                        join(1..531,631..864,954..1133)
                        /gene="NZZ"

```

sporogenesis"
/codon_start=1
/product="NO22LE"
/protein_id="AAD37775.1"
/translation="MATELFPMSTDDNSVGNPDLRLNRLVNVSSGRIETLKSRL
RKPGSKTQCKQKPTLRGMVAKLERQIEEKKQAAATVGTSSVASISNNATRL
PVPDGVVLOGPPLSLGSRNYICGVSGGVWIDPVI SPWGFVETSTHLSLSIN
POMFNASSNNRCDTCKFKKRLDGDONNVRSNGGFSKYMTIPPMNGYDQVLLQSDH
HORSOGFLYDRIARAASVSASSTINIFYNEATNHTGPMEEFGSMEGNPNMGSGGV
KEYEFPKGERVSVVATTSLSLVGDCSPNTIDLSKL"
90
/gene="NZZ"
/note="compared to Arabidopsis thaliana Columbia ecotype
sequence presented in GenBank Accession number AL030978"
/replace="c"
1072
variation
/gene="NZZ"
/note="compared to Arabidopsis thaliana Columbia ecotype
sequence presented in GenBank Accession number AL030978"
/replace="a"
323 a 231 c 255 g 324 t

variation
/gene="NZZ"
/note="compared to Arabidopsis thaliana Columbia ecotype
sequence presented in GenBank Accession number AL030978"
/replace="c"
1072
variation
/gene="NZZ"
/note="compared to Arabidopsis thaliana Columbia ecotype
sequence presented in GenBank Accession number AL030978"
/replace="a"
323 a 231 c 255 g 324 t

Query Match 78.0%; Score 736; DB 8; Length 1133;
Best Local Similarity 83.4%; Pred. No. 1.9e-200;
Matches 944; Conservative 0; Mismatches 0; Indels 188; Gaps 2;
Y 1 TGCGACTTCTCTTCTTCATGTCACAGATCAAAACCTCGGTCGGAACCCCAACGATC 60
b 2 TGCGACTTCTCTTCTTCATGTCACAGATCAAAACCTCGGTCGGAACCCCAACGATC 61
Y 61 TTCTGAGAACACCCGCTTCTTCTGCTCAATAGCTCCGGGAGATCCGGAGACAGACTGA 120
b 62 TTCTGAGAACACCCGCTTCTTCTGCTCAATAGCTCCGGGAGATCCGGAGACAGACTGA 121
Y 121 AGAGTCGTCGCGAACACAGGATCGAAGACAGCTCAGCAAAACAGAGAACCAACGCT 180
b 122 AGAGTCGTCGCGAACACAGGATCGAAGACAGCTCAGCAAAACAGAGAACCAACGCT 181
Y 181 TGAGAGGAATGGGTGTAGCAAAACCTCGAGCTCAGAGATCGAAGAGAAAGAACCAAC 240
b 182 TGAGAGGAATGGGTGTAGCAAAACCTCGAGCTCAGAGATCGAAGAGAAAGAACCAAC 241
Y 241 TCGCCCGCCGACAGTCGGAGACAGCTCATCAGTAGCATGATCTCTTAACACGCTACCC 300
b 242 TCGCCCGCCGACAGTCGGAGACAGCTCATCAGTAGCATGATCTCTTAACACGCTACCC 301
Y 301 GTTTACCCGTACCGGTAGACCCCGGTGTGTGTGTACAGGCTTCCCAAGCTCACTCGGA 360
b 302 GTTTACCCGTACCGGTAGACCCCGGTGTGTGTGTACAGGCTTCCCAAGCTCACTCGGA 361
Y 361 GCAACAGATCTATTGTGTGGAGTCGGGTGGGTGATGATGATGATGATGATGATGAT 420
b 362 GCAACAGATCTATTGTGTGGAGTCGGGTGGGTGATGATGATGATGATGATGATGAT 421
Y 421 CTCATGGGGTTTTGTGTAGACCTCTCTCCACTACTCATGAGCTCTCTTCAATCTCAATC 480
b 422 CTCATGGGGTTTTGTGTAGACCTCTCTCCACTACTCATGAGCTCTCTTCAATCTCAATC 481
Y 481 CTCATGTTTACGTTCTTCCAAATACCTGTCGACACTTCTGCTCA----- 528
b 482 CTCATGTTTACGTTCTTCCAAATACCTGTCGACACTTCTGCTCAAGGTTGTTGT 541
Y 529 ----- 528
b 542 TTTTAAATCGTTTTCATCAACATGATGTATATATATATATAGTTTTTGCACTTGAAAAGTT 601
Y 529 -----AGAGAAAGCTTTGGATGGTGATCAGATAATG 561
b 602 TTGATTTTATGTAATAAACTCGAGAGAAACCTTTGGATGGTGATCAGATAATG 661
Y 562 TAGTTCGATCCACGGTGGTGGATTTTCGAAATACACAATGATTCCTCTCCGATGAACG 621

Db 662 TAGTTCGATCCACAGGTGGTGGATTTTCGAAATACACAATGATTCCTCTCCGATGAACG 721
QY 622 GCTACGATCAGTATCTTCTCAATCAGATCATCATCAGAGGAGCCCAAGSTTTCCTTTATG 681
Db 722 GCTACGATCAGTATCTTCTCAATCAGATCATCATCAGAGGAGCCCAAGSTTTCCTTTATG 781
QY 682 ATCATAGATCGGTAGAGCAGCTTCAAGTTCTGCTTCTAGTACTACTATTAACCTTAT 741
Db 782 ATCATAGATCGGTAGAGCAGCTTCAAGTTCTGCTTCTAGTACTACTATTAACCTTAT 841
QY 742 TCAACGAGGCAACAATCATAC----- 763
Db 842 TCAACGAGGCAACAATCATACGCTACTAGTATAGTCCATTTATTAATACATATATA 901
QY 764 -----GGGACCAAT 772
Db 902 GGTATATATATATACATGTTGATCTTATTTGATTAACTGGTGGTTTGGGACCAAT 961
QY 773 GGAGGAATTTGGGAGCTACATGGAAGAAACCTAGAAATGATCAGAGGTCGTGAAGA 832
Db 962 GGAGGAATTTGGGAGCTACATGGAAGAAACCTAGAAATGATCAGAGGTCGTGAAGA 1021
QY 833 GTACGAGTTTTTCCGGGAAATATGCTGAAAGAGTTTCAGTGGTGGCTACACGTCGTC 892
Db 1022 GTACGAGTTTTTCCGGGAAATATGCTGAAAGAGTTTCAGTGGTGGCTACACGTCGTC 1081
QY 893 ACTCGTAGTGATGTCAGTCCTTAATACCATTTGTTGCTTGAAGCTTTAA 944
Db 1082 ACTCGTAGTGATGTCAGTCCTTAATACCATTTGTTGCTTGAAGCTTTAA 1133

RESULT 3
ATF27G19/c
LOCUS
DEFINITION
Arabidopsis thaliana DNA chromosome 4, BAC clone F27G19 (ESSA project).
ACCESSION
AL078467
VERSION
AL078467.1 GI:4972055
KEYWORDS
SOURCE
Arabidopsis thaliana.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 82381)
Bevan, M., Hilbert, H., Braun, M., Holzer, E., Brandt, A.,
Duesterhoeft, A., Bancroft, I., Mewes, H. W., Mayer, K. F. X., Lemcke, K.,
Mannhaupt, G. and Schueller, C.
Unpublished
JOURNAL
REFERENCE
2 (bases 1 to 82381)
EU Arabidopsis sequencing project.
AUTHORS
Direct Submission
TITLE
Submitted (30-MAY-1999) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
schueller@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
COMMENT
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

FEATURES
Location/Qualifiers
1..82381
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
misc_feature
1..30951
/note="overlap to BAC M4122; please refer to this entry
for analysis and annotation"
misc_feature
5126..5143
/note="there is a mismatch in this homopolimeric stretch

```

to the overlapping BAC M4122; 18(F27G19) versus 17(M4122)
A's; the genomic DNA clearly shows 18 A's"
31517..32602
/gene="F27G19.10"
complement(join(31517..31960,32044..32335,32394..32602))
/gene="F27G19.10"
complement(join(31517..31960,32044..32335,32394..32602))
/gene="F27G19.10"
/note="similarity to Arabidopsis thaliana nap gene,
PID:e1234813
contains EST gb:R83977"
/codon_start=1
/product="putative protein"
/protein_id="CAB43873.1"
/db_xref="GI:4972056"
/translation="MGVREKDPQLAQLSLPPGFRFYPTDELLVOYLCKRVAGYHFSLQ
VIGDILYFDPWDLPSKQCTTFVGEYNVLGKALFGKEKWYFSPDRKYPNGSR
PNRVAGSGYKATGTDKIIITADGRRVGIKKALVYAGKAPKGTKTNIIMHEVRLIHS
RSHGSKLDWLCRIYKKTSCSQROAVTPVOACREHSTNGSSSSSSQDLDVDSF
PEIKDSFNLPRNSRLTILNFDWASLAGLNPPELAPTNGLPLSYGGYDFAAEG
EASGHVNRQNSSLGTQSGYSSSGFGVSGQTTFPRQ"
complement(31517..31960)
/gene="F27G19.10"
/number=1
complement(31961..32043)
/gene="F27G19.10"
/number=1
complement(32044..32335)
/gene="F27G19.10"
/number=2
complement(32336..32393)
/gene="F27G19.10"
/number=2
complement(32394..32602)
/gene="F27G19.10"
/number=3
34968..35014
/note="47bp at tandem repeat"
/gene="F27G19.20"
complement(join(36023..36688,36772..37103,37174..37326,
37369..38011,38273..38386))
/gene="F27G19.20"
complement(join(36023..36688,36772..37103,37174..37326,
37369..38011,38273..38386))
/gene="F27G19.20"
/note="similarity to D.melanogaster P element Casper-1
gene (white protein), PID:g870996
contains ABC transporters family signature
[VSGGRRKRVISGOEI], ATP/GTP-binding site motif A (P-loop)
[GPSGSGKT] [ALTRIGKT]"
/codon_start=1
/product="putative protei"
/protein_id="CAB43874.1"
/db_xref="GI:4972057"
/translation="MDNQVSMVDVETPIAKTNDRLSPFSIFKANNPVTLKFNELVY
TVKLDSOCGPKNDKTEPTILKGLTGIVKPEILAMLGPSGSGKTSLTALGGRVG
EGKGLTGINSYNNKPLSKAVRTTGFVTDALYPLNTVETLFTALLRLPNSEFK
QKIKQAKAVHELGLDKCKDRIIGGFLRGVSGGKRVSGICQELLNPSLFLDEP
TSLGDSLTQRIIVSLIWEIARGRTVVTTHQPSKNPVIYFGLNSAMDYFASVGYSP
LYVERINPSDFLIANGKPLVISCNPSVSGDSQRPAMKAALVAFYKTNLSDVIN
FYVKGODLCKNPKRESSRVATNYGDWPTTWQOFCVLLKRGKQRHDSFGSMKVAQI
FIVSFLGGLLWQTKISRLQDQIGLLFFTSFMAFPPLQQIFTPPOBRAMLOKERS
GWRLSPYELSRVGDLPHELILPTCLFVITYMAGLNHLNLANFEVLLYLVHVIYS
GGLGLAGLAWDQKSATTLGSLVIMTLFLLAGGYVQHVVPFISWIKVISGITYIKL
LILGQITANELYPCDNGKLRCHVDFEGIKRHGFNSGLVLSALAUTAMLVYRVYAI
ALTRIGKTKSG"
complement(36023..36688)
/gene="F27G19.20"
/number=1
complement(36689..36771)
/gene="F27G19.20"
/number=1

```

```

complement(36772..37103)
/gene="F27G19.20"
/number=2
complement(37104..37173)
/gene="F27G19.20"
/number=2
complement(37174..37326)
/gene="F27G19.20"
/number=3
complement(37327..37368)
/gene="F27G19.20"
/number=3
complement(37369..38011)
/gene="F27G19.20"
/number=4
complement(38012..38272)
/gene="F27G19.20"
/number=4
complement(38273..38386)
/gene="F27G19.20"
/number=5
42406..46325
/gene="F27G19.30"
join(42406..42464,42566..42742,42832..42953,43062..43114,
43203..43246,43346..43498,43587..43813,43912..45359,
45432..46325)
/gene="F27G19.30"
/note="contains EST gb:Z37620, Z37621"
/codon_start=1
/product="COPI-interacting protein 7 (CIP7)"
/protein_id="CAB43875.1"
/db_xref="GI:4972058"
/translation="MDPTRLRDLALFQLTPTRCDLVIFSGGENEKLASGIFOPFVT
HLKSVSDQISKGYSVTLRPPSVGVFWTKVTLQRFVFTTPEVLRSVTLKEFIEQ
IEDSIQANAAAJAGRAEGNELGWTQSOKTALSKTGTGDTVEENSKVGLORVLE
NRKAALKEQAMAYARALYVGFELYMDLFSADAFASRLREACVNFVLCRRKNE
DRMWYDQITAMAFRPELTFMGDSGIVLAGENDLNLATNKKHNSMDASQGSFET
GOEGRAMAMPNPQFPQYMQNFQGHGYPFPYFMFGMQGSPYFPHGNMOWPNMGDVE
SNEKSKKKKKKKKKKKKQDESAPESDNGSTSESDGNEGKQSKVYIRNINYI
TSKRNGAKESDSEGEFVDGDSIKQOVEEAGISVERHAKSTSHRKHKSHNGD
DSSSKETKGNNDWAFONLLKONDSPRELLRLSSITALNMASEVVKRPPSDSF
LVAIGNEWGRETSTIEKNAGENVRIIRKNNYDEMLNPGSDERSYSQAEMSVHD
GKLRTNKAEDWFIRNAGPETDPSLVKTFVGDHFLNKKSERDLVDDSFMIHSRV
ENQVDSRLRTEIMDLVYGTQQNSAPENTPHEPDLYMYLGRQDVKPTLLPWT
EIDFNTNLAORTSRIDLITATKASAGQTLDGKEKSRGISKGDASKRASSRPPA
SKAKRPANGSAVSKSEMEERKKRMEELLQORKRIAEKSSGGVSSSLASKT
PVTYSKSSIKNEKTPAAQSKAPVLRSSIERLAVARTAPREPQOKPVKRTSKP
SGYTKAQERKSKKIGSDAKSVLSRDPSELEIKETVEDSHSVLSKQVDALPAVA
SYDDPKDILKELHSLPSEETARKNRPNEIIAKKVQDQTKIDQETVKRYSDEKQIT
TKHYSEDYEQVQAQSKVPKSKSVTFSTNMEEKYYFSPAVSEIDISTPPATEADHS
RKKWNSETS PKATAKAVFRKLLMFORKK"
42406..42464
/gene="F27G19.30"
/number=1
42465..42565
/gene="F27G19.30"
/number=1
42566..42742
/gene="F27G19.30"
/number=2
42743..42831
/gene="F27G19.30"
/number=2
42832..42953
/gene="F27G19.30"
/number=3
42954..43061
/gene="F27G19.30"
/number=3
43062..43114
/gene="F27G19.30"
/number=4
43115..43202

```

/gene="F27G19.30"
/number=4
43203..43246
/gene="F27G19.30"
/number=5
43247..43345
/gene="F27G19.30"
/number=5
43346..43498
/gene="F27G19.30"
/number=6
43499..43586
/gene="F27G19.30"
/number=6
43587..43813

Query Match 77.6%; Score 732.8; DB 8; Length 82381;
Best Local Similarity 83.2%; Pred. No. 1.6e-199;
Matches 942; Conservative 0; Mismatches 2; Indels 188; Gaps 2;

Y 1 TGGGCACTTCTCTCTCTTCATGTCACAGATCAAAACTCCGTCGSAACCCCAACGATC 60
D 7047 TGGGCACTTCTCTCTCTTCATGTCACAGATCAAAACTCCGTCGSAACCCCAACGATC 6988

Y 61 TTCTGAGAAACACCGCTCTTGTGCTCAATAGTCTCGGCGAGATCCGGACAGACACTGA 120
D 6987 TTCTGAGAAACACCGCTCTTGTGCTCAATAGTCTCGGCGAGATCCGGACAGACACTGA 6928

Y 121 AGAGTCGTGTCGGAACACGATCGAACACAGTTCAGCAAAACAGAGAACCAACGT 180
D 6927 AGAGTCGTGTCGGAACACGATCGAACACAGTTCAGCAAAACAGAGAACCAACGT 6868

Y 181 TGAGAGGAATGGTGTAGCAAGCTCGACGTCAGAGATCGAAGAAAGAAAGAACCAAC 240
D 6867 TGAGAGGAATGGTGTAGCAAGCTCGACGTCAGAGATCGAAGAAAGAAAGAACCAAC 6808

Y 241 TCGCCGCCGCCACAGTCGGAGACAGTCATCATAGTAGCATCGATCTCTTCAACACGCTATCC 300
D 6807 TCGCCGCCGCCACAGTCGGAGACAGTCATCATAGTAGCATCGATCTCTTCAACACGCTATCC 6748

Y 301 GTTATCCGTCACGGTAGACCGGGTGTGTCGTACAGCGTCCCAAGCTCACTCGGGA 360
D 6747 GTTATCCGTCACGGTAGACCGGGTGTGTCGTACAGCGTCCCAAGCTCACTCGGGA 6688

Y 361 GCAACAGGATCATGTGTGTGAGTCGGGTCCGGTCAGGTTATGATCGACCCGGTATTT 420
D 6687 GCAACAGGATCATGTGTGTGAGTCGGGTCCGGTCAGGTTATGATCGACCCGGTATTT 6628

Y 421 CTCATGGGGTTTTGTGAGACCTCTCCACTACTCATGAGCTCTCTTCAATCTCAAAATC 480
D 6627 CTCATGGGGTTTTGTGAGACCTCTCCACTACTCATGAGCTCTCTTCAATCTCAAAATC 6568

Y 481 CTCAAATGTTTAAAGCTTCTTCCAATAATCGCTGTGACACTTGTCTCA----- 528
D 6567 CTCAAATGTTTAAAGCTTCTTCCAATAATCGCTGTGACACTTGTCTCAAGGTTGTTGT 6508

Y 529 ----- 528

D 6507 TTTTAACTGTTTTCATCAACATGATGATATATATATATATATATATATATATATATAT 6448

Y 529 -----ACAGAAAGCTTTGGATGGTGATCGATCAAGTAATG 561
D 6447 TTGATTTTATTTATGTAAGAACTGCAGAGAAAGCTTTGGATGGTGATCGATCAAGTAATG 6388

Y 562 TAGTTCGATCCACCGTGGTATTTCGAATACAAATGATTCCTCCCGATGAAGC 621
D 6387 TAGTTCGATCCACCGTGGTATTTCGAATACAAATGATTCCTCCCGATGAAGC 6328

Y 622 GTACGATCAGTATCTTCTCAATCAGATCATCATCAGAGGCCAAGGTTTCTTTATG 681
D 6327 GTACGATCAGTATCTTCTCAATCAGATCATCATCAGAGGCCAAGGTTTCTTTATG 6268

Y 682 ATCATAGAAATCGTAGACGAGCTTCAGTTTCTGCTCTAGTAGTACTACTATTAATCCTTATT 741

Db 6267 ATCATAGATCCCTAGACGAGCTTCAGTTCTGCTTAGTACTACTATTAATCCTATT 6208
Qy 742 TCACGAGGCAACAAATCATAC----- 763
Db 6207 TCACGAGGCAACAAATCATACGGTACTAAGTATAGTCCATTATTAATCACTATATA 6148
Qy 764 -----GGGACCAAT 772
Db 6147 GGTATATATGATATAAAGTGTGATCTTATTTGATTTAAGTGGTGGTTAGGACCAAT 6088
Qy 773 GGAGGAATTTGGGACTACATGGAAGGAACCCCTAGAAATGATCAGGAGTGTGAAGGA 832
Db 6087 GGAGGAATTTGGGACTACATGGAAGGAACCCCTAGAAATGATCAGGAGTGTGAAGGA 6028
Qy 833 GTACGAGTTTTTCCGGGAAATATGTTGAAAGAGTTCAGTGGTGGCTACAAACGTCGTC 892
Db 6027 GTACGAGTTTTTCCGGGAAATATGTTGAAAGAGTTCAGTGGTGGCTACAAACGTCGTC 5968
Qy 893 ACTCGTAGTGATTCGAGTCCTAATACCATGATTGTCCTTGAAGCTTTAA 944
Db 5967 ACTCGTAGTGATTCGAGTCCTAATACCATGATTGTCCTTGAAGCTTTAA 5916

RESULT 4
ATM4122/c
LOCUS
DEFINITION Arabidopsis thaliana DNA 83371 bp DNA linear PLN 10-MAR-2000
project).

ACCESSION AL030978
VERSION AL030978.1 GI:3269280

KEYWORDS
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
EU Arabidopsis sequencing, project.
Direct Submission
Submitted (09-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lencke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

FEATURES
Location/Qualifiers
1..83371
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
540..784
737..3246
/gene="AT4g27200"
join(737, 1811, 1828, 2002, 2037, 3246)
/gene="AT4g27200"
/note="similarity to retrotransposon Opie-2, Zea mays,
PID:g1657768
contains EST gb:T46634, AA404771, AI997526.1"
/codon_start=1
/product="putative protein"
/protein_id="CAA19714.1"
/db_xref="GI:3269281"
/db_xref="SPTREMBL:O81823"
/translation="MLVKKRSTSSWRIGSACFPPLRDYAEKNFNPCLKCVFLGYNE
KYGRCLYPTGLYISRHVIFDESVPFSTYKHLHPQRPPLLAWLKSSSPAP
STSTSPSRSPFLTADPPLPQKPLPLPLVPISSVSHASNITQSPDFDSERT
DLSAGTGDSSHQASDEETQQASNVHOTPASTNVHPMTRKVGSKPNRY
VFLSHKVSYPKTVTAALKHPGTGAMTEIGNCSTQTSWLSVPYKSDMVLGNKW
FRTLKADSTLKLKARIVAKGFLQEGIDYLETYSVPVRTPTVRLVHLATLNWDI
KQMDVKNAPLHGLDKETVYMTQPRANRDHVCLLHKSITYGLKQSPRAWDKFSTPLLEF
GFFCKNSDPISLFIYAHNNLILLISQTLTSLAALNKEFRMTDMGQHSILTLGIQVQ

gene	15169..17649	77.6%;	Score 732.8;	DB 8;	Length 83371;
cds	/gene="AT4g27240"	Join(16169..16829,17015..17649)			
	/gene="AT4g27240"	/note="Contains Zinc finger, C2H2 type, domain AA198-219 contains EST gb:AI993453.1, T88464"			
	/codon_start=1				
Query Match		77.6%;	Score 732.8;	DB 8;	Length 83371;
Best Local Similarity		83.2%;	Pred. No. 1.6e-199;		
Matches 942;	Conservative 0;	Mismatches 2;	Indels 188;	Gaps 2;	
1	Y	Y	Y	Y	Y
59467	b	b	b	b	b
61	Y	Y	Y	Y	Y
59407	b	b	b	b	b
121	Y	Y	Y	Y	Y
59347	b	b	b	b	b
181	Y	Y	Y	Y	Y
59287	b	b	b	b	b
241	Y	Y	Y	Y	Y
59227	b	b	b	b	b
301	Y	Y	Y	Y	Y
59167	b	b	b	b	b
361	Y	Y	Y	Y	Y
59107	b	b	b	b	b
421	Y	Y	Y	Y	Y
59047	b	b	b	b	b
481	Y	Y	Y	Y	Y
58987	b	b	b	b	b
529	Y	Y	Y	Y	Y
58927	b	b	b	b	b
529	Y	Y	Y	Y	Y
58867	b	b	b	b	b
562	Y	Y	Y	Y	Y
58807	b	b	b	b	b
622	Y	Y	Y	Y	Y
58747	b	b	b	b	b
682	Y	Y	Y	Y	Y
58687	b	b	b	b	b
742	Y	Y	Y	Y	Y
58627	b	b	b	b	b
764	Y	Y	Y	Y	Y

Db	58567	GGTATATATGTATATAACTGTTGATCTATTGATTTAACTGGTGGGTTTAGGACCAAT	58508
QY	773	GGAGGAATTTGGGAGCTACATGGAAAGGAAACCCCTAGAAATGGATCAGAGAGTGTCAGGA 832	
Db	58507	GGAGGAATTTGGGAGCTACATGGAAAGGAAACCCCTAGAAATGGATCAGAGAGTGTCAGGA 58448	
QY	833	GTACGAGTTTTTTCCGGGGAAATATGCTGGAAGAGTTTCAGTGTGCTCAACAGTCGTC 892	
Db	58447	GTACGAGTTTTTTCCGGGGAAATATGCTGGAAGAGTTTCAGTGTGCTCAACAGTCGTC 58388	
QY	893	ACTCGTAGGTGATGTCAGTCCCTAATACCAATGATTTCCCTTGAAGCTTAA 944	
Db	58387	ACTCGTAGGTGATGTCAGTCCCTAATACCAATGATTTCCCTTGAAGCTTAA 58336	
RESULT 5			
ATCHRIV66/C			
LOCUS	ATCHRIV66	198372 bp	DNA linear PLN 16-MAR-2000
DEFINITION	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 66.		
ACCESSION	AL161566		
VERSION	AL161566.2	GI:7269538	
KEYWORDS	Arabidopsis thaliana.		
SOURCE	Arabidopsis thaliana		
ORGANISM	Arabidopsis thaliana		
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.			
REFERENCE	1 (bases 1 to 84417)		
AUTHORS	Leclarny,A., Chefor,F., Krivitzky,M., Kreis,M., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.		
JOURNAL	Unpublished		
REFERENCE	2 (bases 83512 to 129491)		
AUTHORS	Robben,J., Grymonprez,B., Volckaert,G., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.		
JOURNAL	Unpublished		
REFERENCE	3 (bases 127847 to 198372)		
AUTHORS	Reichert,B.J., Barel,E., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.		
JOURNAL	Unpublished		
REFERENCE	4 (bases 185358 to 185634)		
AUTHORS	Volckaert,G., Grymonprez,B., Voet,M., Robben,J., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.		
JOURNAL	Unpublished		
REFERENCE	5 (bases 1 to 198372)		
AUTHORS	EU Arabidopsis sequencing,project.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@ebc.ac.uk		
COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ this fragment has an overlap with ATCHRIV65 at the 5' end and an overlap with ATCHRIV67 at the 3' end.		
FEATURES	Location/Qualifiers		
Source	1..198372		
	/organism="Arabidopsis thaliana"		
	/variety="Columbia"		
	/db_xref="taxon:3702"		
	/chromosome="4"		
gene	7146..8548		
	/gene="At4g26860"		
gene	complement(join(7146..7223,7356..7437,7485..7689,7764..7851,7956..8002,8081..8147,8327..8548))		
	/gene="At4g26860"		
CDS	complement(join(7146..7223,7356..7437,7485..7689,7764..7851,7956..8002,8081..8147,8327..8548))		
	/gene="At4g26860"		
/note="strong similarity to Proline synthetase associated - Homo sapiens, PID:d1037830			

Contains Uncharacterized protein family UPF0001 signature
AA81-95
contains EST gb:AI995798.1, N97000, T76579"
/codon_start=1
/product="putative Proline synthetase associated protein"
/protein_id="CA879541.1"
/db_xref="GI:7269539"
/translation="MAAPAVEATVASALRSVILRAKAAEOYGRDAERVRVLPVSKTK
PVSLLRIQIDAGHRCFGENYVQGIIDKAPOLPEDIHWFVGHLSQNKAKTLITGVPNL
AMVGVGSKVANHLDRAVSNLGRHPLKVLVQNTSGEYKSGIEPSSVVELARHVKH
HCPNLVFSGLMTIGMDPYTSTPENRFVYLLIGLYLFVFSYGYKPYLPRADVCRAKG
MAEQFELSMGSGCDFELAIEMGSTNVRVSGSTIFGPPEYPKKTT"
complement(7146..7223)
/gene="AT4g26860"
/number=1
exon
complement(7224..7355)
/gene="AT4g26860"
/number=1
exon
complement(7356..7437)
/gene="AT4g26860"
/number=2
intron
complement(7438..7484)
/gene="AT4g26860"
/number=2
exon
complement(7485..7689)
/gene="AT4g26860"
/number=3
intron
complement(7690..7763)
/gene="AT4g26860"
/number=3
exon
complement(7764..7851)
/gene="AT4g26860"
/number=4
intron
complement(7852..7955)
/gene="AT4g26860"
/number=5
exon
complement(7956..8002)
/gene="AT4g26860"
/number=5
intron
complement(8003..8080)
/gene="AT4g26860"
/number=5
exon
complement(8081..8147)
/gene="AT4g26860"
/number=6
intron
complement(8148..8326)
/gene="AT4g26860"
/number=6
exon
complement(8327..8548)
/gene="AT4g26860"
/number=7
gene
9236..11474
/gene="AT4g26870"
join(9236..9715,9783..9893,9996..10079,10191..10481,
10555..10665,10874..11069,11156..11270,11348..11474)
/gene="AT4g26870"
/note="similarity to aspartate-tRNA ligase (EC 6.1.1.12)
-Methanobacterium thermoautotrophicum, GB:AE000809
Contains Aminoacyl-transfer RNA synthetases class-II
signatures AA293-311; Aminoacyl-transfer RNA synthetases
class-II signatures AA473-482
contains EST gb:Z34062, AA720362"
/codon_start=1
/product="putative aspartate-tRNA ligase"
/protein_id="CA879542.1"
/db_xref="GI:7269540"
/translation="MVGSEVLECEGKISKESKKRAAKLEKLRKQREATSSLS
LEEDSCSSNYGDVTTNELQSAVEGKELTDVSNLVEEIVGSEVSI RGLHKNRLVGT
KLFVILRESGFTVOCVETRYGVANIKFVKOLRESVVELIGVYSHKPKPLTGTTQQ
VETHRKMKFCLSRPLNPLVVEDAARSEDIEKSGKDTLRNNRVLDIRTPANQAIER
IQCVQVIAFREYLOQVGFLEIHTPKLIAGSEGGSAVFRLDYKGQACIAGSPQLHKQ
MAICGDMRREYVGVFRAEDSFTHRLCEFFVGLDVEMEIRMYSEIMDLVGELEPFI
1
MLHRYPSAVRPEYTMPYENDSNYSNSFOVFI RGEITMSGAORIHDPELLEKRAECGI
DYKTTISTVIDAFRYGAPPHGGFGVGLERVMLLCALNNIRKTSLEFRDSQRITP"
9236..9715
/gene="AT4g26870"
/number=1
intron
9716..9782
/gene="AT4g26870"
/number=1
exon
9783..9893
/gene="AT4g26870"
/number=2
intron
9894..9995
/gene="AT4g26870"
/number=2
exon
9996..10079
/gene="AT4g26870"
/number=3
intron
10080..10190
/gene="AT4g26870"
/number=3
exon
10191..10481
/gene="AT4g26870"
/number=4
intron
10482..10554
/gene="AT4g26870"
/number=4
exon
10555..10665
/gene="AT4g26870"
/number=5
intron
10666..10873
/gene="AT4g26870"
/number=5
exon
10874..11069
/gene="AT4g26870"
/number=6
intron
11070..11155
/gene="AT4g26870"
/number=6
exon
11156..11270
/gene="AT4g26870"
/number=7
intron
11271..11347
/gene="AT4g26870"
/number=7
exon
11348..11474
/gene="AT4g26870"
/number=8
gene
12120..12578
/gene="AT4g26880"
complement(12120..12578)
gene
/gene="AT4g26880"
complement(12120..12578)
CDS
/gene="AT4g26880"
/note="similarity to STIG1 protein - common tobacco,
EMBL:X77823"
/codon_start=1
/product="STIG1 like protein"
/protein_id="CA879543.1"
/db_xref="GI:7269541"
/translation="MAFYKLLVSIATITTAITAVITTTNNNTTIREYTSFDPASTPT
IRPNRLLAOKVEYGERNPNAADHCNRPICTPYGGGSGNSTMTCCNNKCIDVSSDDNN
CGACKNKKCFSTCCRCGCVVAVYDKRHCGCNPCELGEFCVYGLCNYA"
complement(12120..12578)
/gene="AT4g26880"
/number=1
gene
15927..17261
/gene="AT4g26890"
15927..17261
/gene="AT4g26890"
/note="similarity to NPK1-related protein kinase 1L (ANP1)
- Arabidopsis thaliana, PID:d1022712
Contains Protein kinases signatures and profile

Query Match 77.6%; Score 732.8; DB 8; Length 198372;
Best Local Similarity 83.2%; Pred. No. 1.6e-199;
Matches 942; Conservative 0; Mismatches 2; Indels 188; Gaps 2;

Y 1 TGGCGACTTCTCTCTTCTTCATGTCACAGATCAAAACTCCGTCGGAACCCCAACAGCATC 60
b 187313 TGGCGACTTCTCTCTTCTTCATGTCACAGATCAAAACTCCGTCGGAACCCCAACAGCATC 187254

Y 61 TTCTGAGAAACACCCCTCTTCTGCGTCAATAGTCTCGGCGAGATCCGGACAGACACTGA 120
b 187253 TTCTGAGAAACACCCCTCTTCTGCGTCAACAGCTCCGCGAGATCCGGACAGACACTGA 187194

Y 121 AGAGTCGTGGTGGAAACACAGGATCAACAGAGTCAGCAAAAACAGAGAAACCAACGT 180
b 187193 AGAGTCGTGGTGGAAACACAGGATCAACAGAGTCAGCAAAAACAGAGAAACCAACGT 187134

Y 181 TGAGAGAAATGGGTGTAGCAAAAGCTCGAGCGTCAGAGATCGAAGAAAGAAAGAACAC 240
b 187133 TGAGAGAAATGGGTGTAGCAAAAGCTCGAGCGTCAGAGATCGAAGAAAGAAAGAACAC 187074

Y 241 TCGCCGCCGCCACAGTCGAGACACGTCATCAGTAGCATCTCTTAACAACGCTACCC 300
b 187073 TCGCCGCCGCCACAGTCGAGACACGTCATCAGTAGCATCTCTTAACAACGCTACCC 187014

Y 301 GTTTACCCGTACGGGTAGACCCGGGTGTGTGCTACAGGCTTCCCAAGCTCACTCGGGA 360
b 187013 GTTTACCCGTACGGGTAGACCCGGGTGTGTGCTACAGGCTTCCCAAGCTCACTCGGGA 186954

Y 361 GCAACAGGATCTATTCTGCTGGAGTCGGGTGGGTGAGTATGATCGACCCGGTTATT 420
b 186953 GCAACAGGATCTATTCTGCTGGAGTCGGGTGGGTGAGTATGATCGACCCGGTTATT 186894

Y 421 CTCATCGGGGTTTGTGTGAGACTCTCCCTCACTACTCATGAGCTCTCTCAATCTCAAAATC 480
b 186893 CTCATCGGGGTTTGTGTGAGACTCTCCCTCACTACTCATGAGCTCTCTCAATCTCAAAATC 186834

Y 481 CTCAAATGTTTAACGTTCTTCCAAATAATCGCTGTGACACTTGCTTCA----- 528
b 186833 CTCAAATGTTTAACGTTCTTCCAAATAATCGCTGTGACACTTGCTTCAAGGTTTGTGTTGT 186774

Y 529 ----- 528
b 186773 TTTTAAATCGTTTTCATCAACATGATGTGATATATATATATATAGTTTTCACCTTGAAAAGTT 186714

Y 529 ----- AGAAGAAAGCTTTGGATGGTATCAGATAATG 561
b 186713 TTGATTTTATTATGTATAAAACTCCAGAGAAAGCTTTGGATGGTATCAGATAATG 186654

Y 562 TAGTTCGATCCAAACGGTGGGATTTTCGAAATACAAATGATTCCTCTCGATGAACG 621
b 186653 TAGTTCGATCCAAACGGTGGGATTTTCGAAATACAAATGATTCCTCTCGATGAACG 186594

Y 622 GCTACGATCAGTATCTTCTCAATCAGATCATCATCAGAGGCCAAGGTTTCCTTTATG 681
b 186593 GCTACGATCAGTATCTTCTCAATCAGATCATCATCAGAGGCCAAGGTTTCCTTTATG 186534

Y 682 ATCATAGATCGCTAGACGAGCTTCAGTTCTCTCTCTAGTACTACTATTAATCCTTATT 741
b 186533 ATCATAGATCGCTAGACGAGCTTCAGTTCTCTCTCTAGTACTACTATTAATCCTTATT 186474

Y 742 TCAACGAGGCAACAAATCATAC----- 763
b 186473 TCAACGAGGCAACAAATCATACCGGTACTAAGTATAGTCCATTATTAACTCATATATA 186414

Y 764 -----GGACCAAT 772
b 186413 GGTATATATGATATAACRTTTCATCTTATTGATTTAATCTGTTGATTAAGGACCAAT 186354

Y 773 GGAGGAATTTGGAGCTACATGGAAGAAACCCCTAGAAATGATCAGGAGGTGTGAAGA 832
b 186353 GGAGGAATTTGGAGCTACATGGAAGAAACCCCTAGAAATGATCAGGAGGTGTGAAGA 186294

Y 833 GTACGAGTTTTCGCGGGAATATATGTTGAAAGAGTTTTCAGTGGTGGCTACACAGCTCGTC 892

Db 186293 GTACGAGTTTTTCCGGGAATATGTTGAAAGAGTTTCAGTGGTGGCTAAACGTCGTC 186234

QY 893 ACTCGTAGGTGATTCAGTCCTTAATACCATTTGATTTGCTCTTGAAGCTTTAA 944

Db 186233 ACTCGTAGGTGATTCAGTCCTTAATACCATTTGATTTGCTCTTGAAGCTTTAA 186182

RESULT 6

LOCUS I66494 7218 bp DNA linear PAT 28-DEC-1997

DEFINITION Sequence 14 from patent US 5670367.

ACCESSION I66494

VERSION I66494.1 GI:2724471

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 7218)

AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.

TITLE Recombinant fowlpox virus

JOURNAL Patent: US 5670367-A 14 23-SEP-1997;

FEATURES

Location/Qualifiers

1..7218

source /organism="unknown"

BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others

ORIGIN

Query Match 4.9%; Score 45.8; DB 6; Length 7218;
Best Local Similarity 8.5%; Pred. No. 0.065;
Matches 17; Conservative 116; Mismatches 68; Indels 0; Gaps 0;

QY 96 GCGGAGATCGGACAGACACACTGAGAGTCTGTGTCGCGAAACAGGATCGAGACAGGT 155.

Db 1236 RRR 1177

QY 156 CAGCAAAACAGAGAACAACCAAGTTGAGAGGAATGGTGTAGCAAAAGCTCGAGCGTCAG 215

Db 1176 RRR 1117

QY 216 AGAATCGAAGAGAAGAAGCAACTCGCGCGCCACATCGGACGACACGTCATCAGTA 275

Db 1116 RRR 1057

QY 276 GCATCATCTCTAACACCGCT 296

Db 1056 CCTCGACCTCGAGCCAGCT 1036

RESULT 7

LOCUS AC104900 141632 bp DNA linear HTG 02-APR-2002

DEFINITION Mus musculus clone RP23-305C1, WORKING DRAFT SEQUENCE, 12 ordered pieces.

ACCESSION AC104900

VERSION AC104900.2 GI:19881874

KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE Mus musculus.

ORGANISM Mus musculus.

REFERENCE 1 (bases 1 to 141632)

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Mus musculus, clone RP23-305C1

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 141632)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N.,

[illegible]


```
/codon_start=1
/transl_table=11
/product="rel-like protein"
/protein_id="CA670609.1"
/db_xref="GI:6782395"
/translation="MAKEINLTGEEVVALAAYMNETDAAFVKA"

BASE COUNT      1728 a      894 c      927 g      1277 t
ORIGIN
      1  ATCAGATCATCATCAGAGGAGCCAGGTTTCCITTATGATCATAGATCGCTAGACGAGC 703
      2  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      3  3058 ATCAGATCTTAAGGAGAAATACAGAAATTTAAAGGAGCATAAAAGTTTGTGCTGAAGCGGG 3117
      4  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      5  704  TTCAGTTTCGCTTCTAGTACTACTATTAATCCTTATTTCAAGCAGGCAACAATCATAC 763
      6  || || || || || || || || || || || || || || || || || || || || ||
      7  3118 TTATGCTAATCTTATTAACATCATCAAGACACATCTGCTGAAAAAGATAAAGATATATC 3177
      8  || || || || || || || || || || || || || || || || || || || || ||
      9  764  GGGACCAATGAGGAATTTGGGAGCTACATGGAAGGAAACCTAGAAATGGATCAGGAGG 823
      10 || || || || || || || || || || || || || || || || || || || || ||
      11  3178 AGATCTAAAGGAGAAATACAGAACTTAAAGGAGCATAAAAGTTTAGCTGAAGCGGTTA 3237
      12 || || || || || || || || || || || || || || || || || || || || ||
      13  824  TGTCAAGGAGTACGAGTTTTCGCGGGAATATGTTGAAGAGTTTCAGTGGTGCTAC 883
      14 || || || || || || || || || || || || || || || || || || || || ||
      15  3238 TGCTAATCTTATTAACATCATCAAGAACAACTTGCTGAAAAAGATAAACTTAATGCTGA 3297
      16 || || || || || || || || || || || || || || || || || || || || ||
      17  884  AA 885
      18 ||
      19  3298 AA 3299

RESULT 13
HSJ393E18
LOCUS
DEFINITION
      1  Human DNA sequence from clone RP3-393E18 on chromosome 6q25.3-26.
      2  Contains an RPL21 (60S Ribosomal protein L21) pseudogene, ESTs and
      3  GSSs, complete sequence.
ACCESSION
      1  AL078604.10 GI:5738668
VERSION
      1  AL078604.10
KEYWORDS
      1  HTG; RPL21.
SOURCE
      1  Homo sapiens.
      2  ORGANISM
      3  Homo sapiens
      4  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      5  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
      6  1 (bases 1 to 116688)
      7  Laird,G.
      8  Direct Submission
      9  Submitted (05-JAN-2000) Sanger Centre, Hinxton, Cambridgeshire,
      10 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
      11 requests: clonerequest@sanger.ac.uk
      12 On Aug 18, 1993 this sequence version replaced gi:5730246.
      13 This sequence has been finished according to sequence map criteria
      14 as follows. An attempt is made to resolve all sequencing problems,
      15 such as compressions and repeats, but not necessarily within known
      16 annotated human repeat sequence elements (e.g. Alu). Where the
      17 sequence is ambiguous, there is an annotation using the 'unsure'
      18 feature key.
      19 This sequence was generated from part of bacterial clone contigs of
      20 human chromosome 6, constructed by the Sanger Centre Chromosome 6
      21 Mapping Group. Further information can be found at
      22 http://www.sanger.ac.uk/HGP/Chr6
      23 During sequence assembly data is compared from overlapping clones.
      24 Where differences are found these are annotated as variations
      25 together with a note of the overlapping clone name. Note that the
      26 variation annotation may not be found in the sequence submission
      27 corresponding to the overlapping clone, as we submit sequences with
      28 only a small overlap as described above.
      29 The following abbreviations are used to associate primary accession
      30 numbers given in the feature table with their source databases:
      31 Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
```

on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep/ RP3-393E18 is
from the library RPL21-3 constructed at the Roswell Park Cancer
Institute by the group of Pieter de Jong. For further details see
http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP3-393E18.

```
FEATURES
      1  Location/Qualifiers
      2  1..116688
      3  /organism="Homo sapiens"
      4  /db_xref="taxon:9606"
      5  /chromosome="6"
      6  /map="q25.3-26"
      7  /clone="RP3-393E18"
      8  /clone_lib="RPL21-3"
      9  189..571
      10 /note="match: GSS: Em:AQ816419"
      11 repeat_region
      12 1613..2047
      13 /note="LTR repeat: matches 1..458 of consensus"
      14 misc_feature
      15 2602..3190
      16 /note="match: GSS: Em:AQ425998"
      17 misc_feature
      18 2602..2924
      19 /note="match: GSS: Em:AQ425996"
      20 misc_feature
      21 3664..4094
      22 /note="match: GSS: Em:AQ082975"
      23 repeat_region
      24 5192..5223
      25 /note="16 copies 2 mer ac 93% conserved"
      26 misc_feature
      27 complement(6061..6545)
      28 /note="match: GSS: Em:AQ570785"
      29 7781..7952
      30 /note="FRAM repeat: matches -119..40 of consensus"
      31 repeat_region
      32 11682..11817
      33 /note="68 copies 2 mer aa 62% conserved"
      34 repeat_region
      35 14918..15974
      36 /note="LTR repeat: matches 5175..6251 of consensus"
      37 repeat_region
      38 16022..16148
      39 /note="LTR repeat: matches 6145..6288 of consensus"
      40 misc_feature
      41 complement(18366..18884)
      42 /note="match: GSS: Em:AQ503651"
      43 repeat_region
      44 19501..19969
      45 /note="LTR repeat: matches 5688..6156 of consensus"
      46 repeat_region
      47 21238..21428
      48 /note="LTR repeat: matches 2401..2585 of consensus"
      49 misc_feature
      50 23524..23792
      51 /note="match: GSS: Em:AQ141736"
      52 repeat_region
      53 23795..24479
      54 /note="LTR repeat: matches 102..783 of consensus"
      55 repeat_region
      56 26344..26925
      57 /note="LTR repeat: matches 1..602 of consensus"
      58 repeat_region
      59 29581..29697
      60 /note="LTR repeat: matches 1015..1133 of consensus"
      61 misc_feature
      62 30416..30805
      63 /note="match: GSS: Em:AQ026748"
      64 repeat_region
      65 31295..31450
      66 /note="LTR repeat: matches 973..1130 of consensus"
      67 misc_feature
      68 complement(31660..31710)
      69 /note="match: GSS: Em:AQ830740"
      70 repeat_region
      71 31791..32278
      72 /note="LTR repeat: matches 139..633 of consensus"
      73 repeat_region
      74 32482..32905
      75 /note="LTR repeat: matches 0..457 of consensus"
      76 repeat_region
      77 40380..40612
      78 /note="LTR repeat: matches 1..222 of consensus"
      79 repeat_region
      80 40655..41009
      81 /note="LTR repeat: matches 348..760 of consensus"
      82 repeat_region
      83 41321..41386
      84 /note="LTR repeat: matches 760..817 of consensus"
      85 repeat_region
      86 42944..43199
      87 /note="LTR repeat: matches 1760..2060 of consensus"
      88 repeat_region
      89 43332..43440
      90 /note="LTR repeat: matches 371..478 of consensus"
      91 misc_feature
      92 44586..44936
      93 /note="match: GSS: Em:AQ232088"
```

```
repeat_region 45171..45224
repeat_region /note="27 copies 2 mer ac 75% conserved"
repeat_region 48492..50137
repeat_region /note="L1M1 repeat: matches 2403..3693 of consensus"
repeat_region 50142..50273
repeat_region /note="66 copies 2 mer aa 71% conserved"
repeat_region 50309..51878
repeat_region /note="L1MA3 repeat: matches 4238..5785 of consensus"
repeat_region 54464..54497
repeat_region /note="17 copies 2 mer aa 91% conserved"
repeat_region 54500..54982
repeat_region /note="L1MA3 repeat: matches 5771..6300 of consensus"
misc_feature 55584..55943
misc_feature /note="match: GSS: Em:AQ280311"
repeat_region 57529..57961
repeat_region /note="L1M1 repeat: matches 5899..6332 of consensus"
repeat_region 58190..58577
repeat_region /note="MSTB repeat: matches 1..426 of consensus"
misc_feature complement(60686..61144)
misc_feature /note="match: GSS: Em:AQ307298"
misc_feature 61511..61896
misc_feature /note="match: GSS: Em:AQ060016"
misc_feature 62945..63226
misc_feature /note="match: GSS: Em:B48326"
repeat_region 63168..63213
repeat_region /note="L1MA8 repeat: matches 6226..6271 of consensus"
repeat_region 63312..63475
repeat_region /note="82 copies 2 mer tt 65% conserved"
repeat_region 64638..64945
repeat_region /note="MSTD repeat: matches 1..394 of consensus"
misc_feature 65353..65308
misc_feature /note="match: GSS: Em:B64542"
repeat_region 73223..73290
misc_feature /note="34 copies 2 mer ac 85% conserved"
misc_feature complement(75094..75626)
misc_feature /note="match: GSS: Em:AQ262078"
misc_feature complement(75114..75596)
misc_feature /note="match: GSS: Em:AQ358217"
misc_feature complement(75117..75627)
misc_feature /note="match: GSS: Em:AQ395001; match: STS: Em:U27136"
misc_feature 75117..75481
misc_feature /note="match: GSS: Em:AQ469918"
misc_feature 75117..75472
misc_feature /note="match: GSS: Em:AQ044426"
misc_feature complement(75132..75621)
misc_feature /note="match: GSS: Em:AQ356505"
misc_feature complement(75136..75621)
misc_feature /note="match: GSS: Em:AQ544386"
misc_feature complement(75152..75627)
gene /note="match: GSS: Em:AQ121360"
75156..75636
CDS /gene="dJ393E18.1"
75156..75636
/gene="dJ393E18.1"
/note="dJ393E18.1 (RPL21 (60S ribosomal protein L21)
pseudogene); match: cDNAs: Em:X15216 Em:U93863 Em:AF086172
pseudogene); match: cDNAs: Em:X89401 Em:D28406; match: ESTs:
Em:M27905 Em:U14967 Em:X89401 Em:D28406; match: ESTs:
Em:A720532 Em:AA312438 Em:AA316830 Em:AA363163
Em:A1224420 Em:AA395400 Em:AA308138 Em:AA192137
Em:AA314735 Em:R82409 Em:AA592053 Em:AA317644 Em:AA313649
Em:AA814644 Em:A1580034 Em:AA215958 Em:A075657
Em:AA6524917 Em:AA296680 Em:AA224869 Em:A1434282
Em:A1630267 Em:AA649672 Em:AA664783 Em:AA807797 Em:C76084
Em:F21489 Em:AA15307 Em:AA843200 Em:AA600830 Em:AA552504
Em:AA225282 Em:AA156225 Em:T59303 Em:A1025106 Em:AA308543
Em:A1025109 Em:AA155735 Em:AA827406 Em:F22300 Em:AA507197
Em:AA652098 Em:AA352569 Em:AA649084 Em:A1685401
Em:AA524967 Em:A1185814 Em:A1066519 Em:AA864658
Em:AA935499 Em:AA148052 Em:A5533361 Em:AA156154
Em:AA528251 Em:A1601216 Em:A1224080 Em:AA147607
Em:AA188054 Em:AA126639 Em:AA854343 Em:AA630928
Em:AA877053 Em:AA554774 Em:AA467819 Em:AA725689
Em:AA1569783 Em:AA316140 Em:AA402451 Em:AA314245
```


Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zemбек, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (13-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 180895)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarato, J., Campoliano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hago, B., Horton, L., Hulme, W., Ilijev, I., Johnson, R., Jones, C., Kanat, A., Karats, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J.J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliou, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zemбек, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (15-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 15, 2002 this sequence version replaced g1:20531913.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information -----
 Center project name: L22539
 Center clone name: 247_L7
 ----- Summary Statistics -----
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 179833 bases at least Q40
 Consensus quality: 179863 bases at least Q30
 Consensus quality: 180141 bases at least Q20
 Insert size: 182000; agarose-fp

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 11367: contig of 11367 bp in length
*
* 11368 11467: gap of 100 bp
*
* 11468 12993: contig of 1526 bp in length
*
* 12994 13093: gap of 100 bp
*
* 13094 16385: contig of 3292 bp in length
*
* 16386 16485: gap of 100 bp
*
* 16486 47027: contig of 30542 bp in length
*
* 47028 47127: gap of 100 bp
*
* 47128 107763: contig of 60636 bp in length

```

```

* 107764 107863: gap of 100 bp
* 107864 145688: contig of 37825 bp in length
* 145689 145788: gap of 100 bp
* 145789 180895: contig of 35107 bp in length.
FEATURES
    source
        1. .180895
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="RP23-247L7"
            /clone_lib="RPCI-23 Female Mouse BAC"
            1. .11367
                /note="assembly_fragment"
                clone_end:SP6
                vector_side:left"
                    misc_feature
                        11468. .12993
                            /note="assembly_fragment"
                            13094. .16385
                                /note="assembly_fragment"
                                16486. .47027
                                    /note="assembly_fragment"
                                    47128. .107763
                                        /note="assembly_fragment"
                                        107864. .145688
                                            /note="assembly_fragment"
                                            145789. .180895
                                                /note="assembly_fragment"
                                                clone_end:T7
                                                    vector_side:right"

```

```

BASE COUNT  54612 a 36242 c 35838 g 53603 t 600 others
ORIGIN

```

```

Query Match      4.38; Score 40.2; DB 2; Length 180895;
Best local Similarity 57.6%; Pred. No. 2.6;
Matches 72; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 115 CACTGAAGAGTCGTGGTGGAAACCCAGGATCGAGACAGGTGACGAAACACAGAGAAAC 174
      ||||| ||| || ||||| || || || || || || || || || || || || ||
Db 24889 CACTGAGGAGAAATGTTTAAAGAGGAGAGACAAAGAGGAGGAGGAGGAGGAGGAG 24830
      || || ||||| || || ||||| || || ||||| ||||| ||||| |||||

QY 175 CAACGTTGAGAGGAATGGGTGTAGCAAAGCTCGAGCGTCAGAGAATCGAAGAGAGAAAGA 234
      || || ||||| || || ||||| || || ||||| ||||| ||||| |||||
Db 24829 GAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 24770
      || || ||||| || || ||||| || || ||||| ||||| ||||| |||||

QY 235 AGCAA 239
      | ||
Db 24769 AAGAA 24765

```

Search completed: January 25, 2003, 23:45:15
 Job time : 4108 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

M nucleic - nucleic search, using sw model

un on: January 25, 2003, 22:29:57 ; Search time 2180 seconds
(without alignments)
7013.092 Million cell updates/sec

title: US-09-701-023-1_COPX_81_1024
erfect score: 944
sequence: 1 tggcgactctctctcttc.....attgtccttgaagcttta 944

coring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

searched: 16154066 seqs, 8097743376 residues

total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 100%
Listing first 45 summaries

post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hic: *
9: gb_est1: *
10: gb_est2: *
11: gb_hic: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rod: *

C	7	40.8	4.3	937	17	CNS006XP	AL060056 Drosophila
	8	40.6	4.3	863	17	BH154230	BH154230 ENTR245TF
	9	40.4	4.3	552	17	AQ338133	AQ338133 HS 5011B
	10	40.4	4.3	859	17	CNS06GK	AL392490 T3 end of
C	11	39	4.1	519	14	BQ515341	BQ515341 EST622756
	12	39	4.1	522	13	B1432782	B1432782 EST535543
	13	39	4.1	525	14	BQ515340	BQ515340 EST622755
C	14	39	4.1	1105	17	CNS07AR0	AL436930 T3 end of
	15	39	4.1	1167	12	BG856706	BG856706 1024047G0
C	16	38.8	4.1	638	14	BQ390257	BQ390257 NISC mq12
	17	38.8	4.1	673	9	AL647711	AL647711 AL647711
	18	38.6	4.1	937	17	CNS006ST	AL065880 Drosophila
C	19	38.4	4.1	637	17	AZ097834	AZ097834 RPCI-23-4
	20	38.4	4.1	653	13	B1927000	B1927000 EST546889
	21	38.2	4.0	477	13	B1466529	B1466529 ie23h04.Y
	22	38	4.0	487	12	BF097623	BF097623 EST415596
	23	37.8	4.0	565	17	AQ490219	AQ490219 RPCI-11-2
C	24	37.4	4.0	331	12	BF587816	BF587816 FMI_33.H1
	25	37.4	4.0	921	17	CNS04NC0	AL298521 Tetraodon
	26	37.2	3.9	544	9	AL630678	AL630678 AL630678
	27	37.2	3.9	551	14	BQ418553	BQ418553 ik46b06.Y
C	28	37.2	3.9	889	17	B21212	B21212 T32J11-Sp6
	29	37.2	3.9	997	17	CNS005TE	AL060767 Drosophila
C	30	37	3.9	660	17	AG142495	AG142495 Pan trogl
	31	37	3.9	695	17	A2938420	A2938420 2M0197C06
	32	36.8	3.9	275	17	A2768950	A2768950 1M0569M08
C	33	36.8	3.9	609	13	B1514873	B1514873 BB160016B
	34	36.8	3.9	636	10	BB635500	BB635500 BB635500
	35	36.8	3.9	902	13	B1856118	B1856118 603382821
	36	36.6	3.9	443	17	AQ850356	AQ850356 LMAJFV1.1
C	37	36.6	3.9	550	17	AQ841718	AQ841718 T137820b
	38	36.6	3.9	609	17	AQ847422	AQ847422 LMAJFV1.1
	39	36.6	3.9	688	17	AZ775117	AZ775117 2M0007108
	40	36.6	3.9	1578	12	BF575807	BF575807 602134848
	41	36.4	3.9	238	17	A2663790	A2663790 RPCI-23-4
	42	36.4	3.9	374	12	BE762203	BE762203 NXC1.083
C	43	36.4	3.9	503	17	BH077388	BH077388 RPCI-24-3
	44	36.4	3.9	665	17	AQ630761	AQ630761 RPCI-11-4
C	45	36.2	3.8	684	17	AG040879	AG040879 Pan trogl

ALIGNMENTS

RESULT 1

B98482/c
LOCUS B98482 592 bp DNA linear GSS 31-MAR-1998
DEFINITION T25H20TRB TAMU Arabidopsis thaliana genomic clone T25H20, DNA sequence.
ACCESSION B98482
VERSION B98482.1 GI:3000561
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 592)
AUTHORS Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter,J.C.
TITLE A BAC End Sequence Database for Identifying Minimal Overlaps in Arabidopsis Genomic Sequencing. Update 3
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: T25H20TRB T25H20TRB T25H20TRB
Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@tigr.org
Seq primer: M13 Reverse

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Match	Length	ID	Description
C 1	312.4	33.1	592	17	B98482 T25H20TRB T
C 2	296.8	31.4	508	17	B67977 T25H20TR TA
C 3	182.2	19.3	641	17	BH472044 BOHUA43TR
C 4	59.4	6.3	775	17	BH472100 BOHUR55TR
C 5	45	4.8	645	10	BE188445 PVB 292 L
C 6	41	4.3	1672	10	BE622474 601441274

TITLE		A BAC End Sequence Database for Identifying Minimal Overlaps in Arabidopsis Genomic Sequencing. Update 3	
JOURNAL		Unpublished (1997)	
COMMENT		Contact: Steve Rounsley Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: rounsley@tigr.org Seq primer: M13 Reverse Class: BAC ends High quality sequence stop: 508.	
FEATURES		Location/Qualifiers	
source		1..508	
		/organism="Arabidopsis thaliana" /strain="Columbia" /db_xref="taxon:3702" /clone="T25H20" /clone_lib="TAMU" /sex="hermaphrodite" /note="Vector: BelobACII; Site_1: HindIII; Site_2: HindIII" ; Produced by Rod Wing"	
BASE COUNT		158 a 117 c 82 g 151 t	
ORIGIN			
Query Match		31.4%; Score 296.8; DB 17; Length 508;	
Best Local Similarity		81.6%; Pred. No. 4.1e-77;	
Matches 408; Conservative 0; Mismatches 2; Indels 90; Gaps 2;			
QY	531	AAGAAACGTTGGATGGTCATCAGAAATAGTTCGATCCACACGGTGGTGGATTTCG	590
Db	508	AAGAACGTTGGATGGTGGATCAGAAATAGTTCGATCCACACGGTGGTGGATTTCG	449
QY	591	AAATACAAATGATTCCTCCGATGAACGGCTACGATCAGTATCTTCTCAATCAGAT	650
Db	448	AAATACAAATGATTCCTCCGATGAACGGCTACGATCAGTATCTTCTCAATCAGAT	389
QY	651	CATCATCAGAGGACCAAGTTTCCTTATGATCATAGAAATCGTAGACGCTTCAGTT	710
Db	388	CATCATCAGAGGACCAAGTTTCCTTATGATCATAGAAATCGTAGACGCTTCAGTT	329
QY	711	TCTGCTTCTAGTACTACTTAAATCCTTATTTCAACGAGGCAACAAATATAC	763
Db	328	TCTGCTTCTAGTACTACTTAAATCCTTATTTCAACGAGGCAACAAATATAC	269
QY	764	-----	763
Db	268	AGTATAGTCCATTATTAATACATATATAGGTATATATCTATATAACGTTGATCTTA	209
QY	764	-----GGACCAATGGAGGAATTTGGAGCTACATGSAAGAA	801
Db	208	TTTGATTTAACTGGTGGGTTTAGGGACCAATGGAGGAATTTGGAGCTACATGSAAGAA	149
QY	802	ACCTAGAAATCGATCAGGAGGTGTGAAGGAGTACGAGTCTTTT-CCGGCAATATGTT	860
Db	148	ACCTAGAAATGGATCAGGAGGTGTGAAGGAGTACGAGTCTTTTACCGGGGAATATGTT	89
QY	861	GAAAGAGTTTCAGTGGGTGCTACAACTGCTAGGTGATTCAGTCTCTTAATACC	920
Db	88	GAAAGAGTTTCAGTGGGTGCTAAAACGCTGCTAGGTGATTCAGTCTCTTAATACC	29
QY	921	ATTGATTTGCTCTTGAAGCT 940	
Db	28	ATTGATTTGCTCTTGAAGCT 9	
RESULT 3			
BH472044/c			
LOCUS			
DEFINITION BOHUA43TR BOHU Brassica oleracea genomic clone BOHUA43, DNA sequence			
641 bp DNA linear GSS 13-DEC-2001			

VERSION BH472044.1 GI:17680155
LOCUS
DEFINITION Brassica oleracea.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 641)
AUTHORS Town.C.D., Van Aken.S., Utterback.T. and Fraser.C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOHUA43TF
Contact: Chris Town
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
FEATURES
source
1..641
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOHUA43"
/clone_lib="BOHU"
/note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 155 a 169 c 140 g 177 t
RIGIN
Query Match 19.3%; Score 182.2; DB 17; Length 641;
Best Local Similarity 68.6%; Pred. No. 6.5e-43;
Matches 367; Conservative 0; Mismatches 108; Indels 60; Gaps 6;
Y 46 GAACCCAAACGATCTCTGAGAAACACCGCTTCTGCTCAATAGCTCGCGGAGATCC 105
|||||
b 641 GAACCCAAACGATCTCTGAGAAACACCTTACCAGCTGTTCAC-----CGCGGAGATCC 588
Y 106 GGACAGAGACACTGA---AGAGTCGTGGTTCGGAACACGAGTCAAGACAGGTTCAGCAA 162
|||||
b 587 GGACAGAGACACCGAAGAGAGCGGTGGTTCGGAACCTGATCGAAGACAGGTTCAGCAA 528
Y 163 AACAGAGAACCAAGCTTGAGAGGATGGGTGTAGCAAGCTCGAGCGTCAGAGATCG 222
|||||
b 527 ACCAGAGAACCGAGCTGAGAGGAATGGGTGTGCTAAGCTCGAGCGAGTACAGCTG 468
Y 223 AAGAGAAAGAACCACTCGCCGCCACACAGCTCGAGACAGTCTATCAGTACATCGA 282
|||||
b 467 AGAAGAGAAAGAAAGAACCGTCTCGTGGAGGAGAGGAGACAGCTCAGCTGC----- 413
Y 283 TCTCTAACACGCTACCGGTTTACCGGTACCGGTAGACCGGGTGTGTGCTACAAAGGCT 342
|||||
b 412 -GAGCCTTACGTCACCGGTTTACCGGTGTACCGGACCGGGGTGTGTGCTCAGAGGCT 384
Y 343 TCCCAAGCTCACTCG-----GGAGCAACA 366
b 353 TTCCAAGCTACGGTGTGGACACACACACAGCTCTCTCGAGGGTATATCTCGAAGCAGGT 294
Y 367 GGATCTATTGTGGAGTCGGGTTCGGGTGAGTATGATCGACCGCGTTATTTCTCCAT 426
b 293 TCTTTCCGGGGAGGAGCGGGTTCGGGTGAGTATGATAGACCGGGTGTGTCTCCTT 234
Y 427 GGGGTTTGTGAGACCTCTCCACTACTCATGAGCTCTTCAATCTCAATCTCAAA 486
b 233 GGGGTTTGTGAGATCC-----ACTCATGAGCTCTTCAATCTCCAAATCTCAAA 180
Y 487 TGTTAACGCTTCTTCAATAATCGGTGTGACACTTGTCTCAAGAGAAAGGTTT 541
b 179 TGTATAAG---CCTCCATATATCTGTGACACTTGTCTCAAGGTTTATTTT 128

RESULT 4
BH472100/c
LOCUS
DEFINITION BH472100 BOHI Brassica oleracea genomic clone BOHIR55, DNA
sequence.
ACCESSION BH472100
VERSION BH472100.1 GI:17680211
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 775)
AUTHORS Town.C.D., Van Aken.S., Utterback.T. and Fraser.C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOHIR55TF
Contact: Chris Town
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
FEATURES
Location/Qualifiers
1..775
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOHIR55"
/clone_lib="BOHI"
/note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 265 a 139 c 135 g 236 t
ORIGIN
Query Match 6.3%; Score 59.4; DB 17; Length 775;
Best Local Similarity 62.2%; Pred. No. 2.8e-06;
Matches 115; Conservative 0; Mismatches 61; Indels 9; Gaps 1;
QY 760 ATACGGGACCAATGGAGGAATTTGGAGCTACATGGAAGAAACCCCTAGAAATGGATCAG 819
|||||
Db 349 ATTAGGCTCAATGGAGGAACCTTGGAGC-----GGAGCCCTAGAAACGGAACA 299
QY 820 GAGGTCTGAGGAGTACGAGTCTTTTCCGGGAAATATGCTGAAGAGTTTCAGTGGTGG 879
|||||
Db 298 GAGATGTGAGGAGTACCGAGTCTTCCCTGGAAGCGATGTAATAATCGGTTTCATCAG 239
QY 880 CTACAACGTCGTCACCTCGTAGTTCAGTTCCTTAATACCATGATTTGCTCTTGAAGC 939
|||||
Db 238 TGCTACATCAGTTGGTGTGATTCAGTCTCCGACATCCACCATCGATTTGCTCTGAGC 179
QY 940 TTTAA 944
|||||
Db 178 TTTAA 174
RESULT 5
BE188445/c
LOCUS
DEFINITION BE188445 645 bp mRNA linear EST 12-MAR-2001
PVB 292 L95-22 Litopenaeus vannamei cdNA similar to Penaeidin 2,
mRNA sequence.
ACCESSION BE188445
VERSION BE188445.1 GI:8888645
KEYWORDS EST.
SOURCE Pacific white shrimp.
ORGANISM Litopenaeus vannamei
Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;

Query Match 4.1%; Score 39; DB 14; Length 519;
Best Local Similarity 54.5%; Pred. No. 2.8;
Matches 78; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Y 548 TGATCAGATAATGTAGTTCGATCCACCGTGGTGGATTTTCGAATACACATGATTC 607
b 359 TTATACAAATATCCACCAACAGTGGTGGTATTAATAATTTCCCAATATACC 300

Y 608 TCCCTCGATGACGGCTAGCATCTCTTCAATCAGATCATCATCAGAGGCCA 667
b 299 TCACCAATAATGGCTATGTAATGGTTAGTGCCACCACCTCTCTGATCCATATACC 240

Y 668 AGGTTTCCTTTATGATCATAGAA 690
b 239 TTGGTGGCCTTATCATATAGAA 217

RESULT 12
LOCUS 1432782
DEFINITION B1432782 522 bp mRNA linear EST 21-AUG-2001
ACCESSION PP3AV35 5' sequence, mRNA sequence.
VERSION B1432782
KEYWORDS B1432782.1 GI:15257472
SOURCE EST.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 522)
AUTHORS Restrepo,S., Griffiths,H.M., Smart,C.D., Cho,J., Chiemingo,A.,
Bouiri,O., Buell,C.R., Ronning,C.M., Fry,W.E. and Baker,B.
TITLE Generation of ESTs from Potato Leaves Challenged with Phytophthora
infestans, Compatible Interaction
JOURNAL Unpublished (2000)
COMMENT Contact: Cathy Ronning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13F-R.

FEATURES
source Location/Qualifiers
1..522
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="PP3AV35"
/clone_lib="P. infestans-challenged leaf"
/tissue_type="leaf"
/dev_stage="6 week old"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; whole plants were challenged with 450,000
sporangia/ml P. infestans US-1(US 940501) in Biotron
(Madison, Wisconsin). Leaf tissue was collected at 1, 2,
5, 12, and 24 hours post-challenge and frozen in liquid
nitrogen immediately upon removal. Kennebec plants showed
no signs of HR. Katahdin plants (susceptible to P.
infestans US-1) were used as controls and showed
infection. NOTE: We cannot exclude the possibility that
this sequence is actually derived from Phytophthora rather
than potato."

BASE COUNT 139 a 128 c 63 g 192 t
ORIGIN

Query Match 4.1%; Score 39; DB 13; Length 522;
Best Local Similarity 54.5%; Pred. No. 2.8;
Matches 78; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Y 548 TGATCAGATAATGTAGTTCGATCCACCGTGGTGGATTTTCGAATACACATGATTC 607
b 161 TTATACAAATATCCACCAACAGTGGTGGTATTAATAATTTCCCAATATACC 220

QY 608 TCCCTCGATGACGGCTAGCATCTCTTCAATCAGATCATCATCAGAGGCCA 667
Db 221 TCACCAATAATGGCTATGTAATGGTTAGTGCCACCACCTCTCTGATCCATATACC 280

QY 668 AGGTTTCCTTTATGATCATAGAA 690
Db 281 TTGGTGGCCTTATCATATAGAA 303

RESULT 13
LOCUS BQ515340
DEFINITION BQ515340 525 bp mRNA linear EST 10-JUN-2002
ACCESSION BQ515340
VERSION BQ515340.1 GI:21374209
KEYWORDS EST.
SOURCE EST.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 525)
AUTHORS Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,
Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and
Karamycheva,S.A.
TITLE Generation of a set of potato cDNA clones for microarray analyses
JOURNAL Unpublished (2002)
COMMENT Other_ESTs: EST622756
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato@tigr.org
This clone is available through the Research Genetics, contact the
Research Genetics for further information 1-800-711-6195 or
cdna@resgen.com
Seq primer: T3.

FEATURES
source Location/Qualifiers
1..525
/organism="Solanum tuberosum"
/cultivar="Kennebec or Binjite"
/db_xref="taxon:4113"
/clone="STM1R50"
/clone_lib="Generation of a set of potato cDNA clones for
microarray analyses mixed potato tissues"
/tissue_type="mixed tissues"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Combination of untreated and Phytophthora
infestans-treated libraries of stolons, leaves, leaflets,
axillary buds of stem explants, petioles, germinating eyes
, tubers, or roots."

BASE COUNT 165 a 120 c 99 g 141 t
ORIGIN

Query Match 4.1%; Score 39; DB 14; Length 525;
Best Local Similarity 54.5%; Pred. No. 2.8;
Matches 78; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 548 TGATCAGATAATGTAGTTCGATCCACCGTGGTGGATTTTCGAATACACATGATTC 607
Db 251 TTATACAAATATCCACCAACAGTGGTGGTATTAATAATTTCCCAATATACC 310

QY 608 TCCCTCGATGACGGCTAGCATCTCTTCAATCAGATCATCATCAGAGGCCA 667
Db 311 TCACCAATAATGGCTATGTAATGGTTAGTGCCACCACCTCTCTGATCCATATACC 370

QY 668 AGGTTTCCTTTATGATCATAGAA 690
Db 371 TTGGTGGCCTTATCATATAGAA 393

1000

This Page Blank (uspto)

M protein - protein search, using sw model

on: January 25, 2003, 23:45:22 ; Search time 81 Seconds

(without alignments)

798.751 Million cell updates/sec

itle: US-09-701-023-4

erfect score: 1634

equence: 1 MATSLFMSIDQNSVGNPD.....TTSSLVGDGSPNTIDLSLKL 314

oring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 671580 seqs, 206047115 residues

total number of hits satisfying chosen parameters: 671580

imum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

SPTREMBL_21.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1634	100.0	314	10 Q9S7B8	Q9S7B8 arabidopsis
2	1628	99.6	314	10 Q81836	Q81836 arabidopsis
3	131	8.0	320	10 Q9LWF1	Q9LWF1 oryza sativ
4	109	6.7	554	3 Q08887	Q08887 saccharomyc
5	104	6.4	575	5 Q9VJX4	Q9VJX4 drosophila
6	103.5	6.3	1082	11 Q9EP02	Q9EP02 rattus norv
7	103	6.3	276	10 Q9SBF4	Q9SBF4 arabidopsis
8	103	6.3	1693	5 Q9VXV8	Q9VXV8 drosophila
9	102	6.2	406	5 Q8T0H7	Q8T0H7 drosophila
10	102	6.2	631	5 Q9V406	Q9V406 drosophila
11	102	6.2	1406	5 Q9VJ42	Q9VJ42 drosophila
12	101.5	6.2	729	5 Q9VCX1	Q9VCX1 drosophila
13	101	6.2	883	10 Q8S218	Q8S218 oryza sativ
14	100	6.1	369	10 Q65249	Q65249 arabidopsis
15	100	6.1	767	17 Q96X10	Q96X10 sulfolobus
16	100	6.1	1059	5 Q9VW51	Q9VW51 drosophila

17	100	6.1	11721	4 Q8WXI7	Q8WXI7 homo sapien
18	99.5	6.1	1034	5 Q20398	Q20398 caenorhabdi
19	99	6.1	352	10 Q22685	Q22685 arabidopsis
20	99	6.1	1478	17 Q8TLD5	Q8TLD5 methanosarc
21	98.5	6.0	607	5 Q3VBP2	Q3VBP2 drosophila
22	98.5	6.0	1323	5 Q3NHX4	Q3NHX4 drosophila
23	98.5	6.0	1376	5 Q3BMZ9	Q3BMZ9 drosophila
24	98.5	6.0	2344	5 Q3N3Y8	Q3N3Y8 caenorhabdi
25	98	6.0	603	5 Q3VW86	Q3VW86 drosophila
26	98	6.0	1126	5 Q8T278	Q8T278 dictyosteli
27	98	6.0	1480	10 Q9SB63	Q9SB63 arabidopsis
28	97.5	6.0	674	5 Q22578	Q22578 caenorhabdi
29	97.5	6.0	690	10 Q940D0	Q940D0 arabidopsis
30	97.5	6.0	698	5 Q44447	Q44447 caenorhabdi
31	97.5	6.0	1321	5 Q962D1	Q962D1 drosophila
32	97	5.9	593	10 Q9FRL7	Q9FRL7 arabidopsis
33	97	5.9	1000	17 Q8TNA9	Q8TNA9 methanosarc
34	96.5	5.9	472	13 Q8UWC8	Q8UWC8 gallus gall
35	96.5	5.9	669	10 Q9ZWK0	Q9ZWK0 arabidopsis
36	96.5	5.9	704	5 Q3VSR3	Q3VSR3 drosophila
37	96	5.9	751	5 Q9I7S7	Q9I7S7 drosophila
38	96	5.9	1408	5 Q24341	Q24341 drosophila
39	95.5	5.8	342	16 Q929A0	Q929A0 listeria in
40	95.5	5.8	690	10 Q9LIB7	Q9LIB7 arabidopsis
41	95.5	5.8	697	5 Q3VZX2	Q3VZX2 drosophila
42	95.5	5.8	894	3 Q03833	Q03833 saccharomyc
43	95.5	5.8	1022	5 Q22308	Q22308 caenorhabdi
44	95.5	5.8	1881	17 Q8TJ57	Q8TJ57 methanosarc
45	95.5	5.8	2232	5 P91365	P91365 caenorhabdi

ALIGNMENTS

RESULT 1

ID	Q9S7B8	PRELIMINARY;	PRT;	314 AA.
AC	Q9S7B8;			
DC	01-MAY-2000 (TRENBLrel. 13, Created)			
DT	01-MAY-2000 (TRENBLrel. 13, Last sequence update)			
DT	01-DEC-2001 (TRENBLrel. 19, Last annotation update)			
DE	SPOROXYTELESS.			
GN	SPL OR NZZ.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. LANDSBERG ERCTA;			
RX	MEDLINE=99396716; PubMed=10465788;			
RA	Yang W.C., Ye D., Xu J., Sundaresan V.;			
RT	"The SPOROXYTELESS gene of arabidopsis is required for initiation of			
RT	sporogenesis and encodes a novel nuclear protein.";			
RL	Genes Dev. 13:2108-2117(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. LANDSBERG ERCTA;			
RX	MEDLINE=99432290; PubMed=10500234;			
RA	Schiefthaler U., Balasubramanian S., Sieber P., Chevalier D.,			
RA	Wisman E., Schneitz K.;			
RT	"Molecular analysis of NOZZLE, a gene involved in pattern formation			
RT	and early sporogenesis during sex organ development in Arabidopsis			
RL	thaliana.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 96:11664-11669(1999).			
DR	EMBL; AF159255; AAD45344.1; -.			
DR	EMBL; AF146794; AAD3775.1; -.			
DR	TRANSFAC; T04440; -.			
SQ	SEQUENCE 314 AA; 34051 MW; B0236D87699ACEC3 CRC64;			

Query Match 100.0%; Score 1634; DB 10; Length 314;
Best Local Similarity 100.0%; Pred. No. 1,le-130;

```
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATSLFFMSTQNSVGNPNLLNRLVNVSSGEIRTEITLKSGRKPGSKTGQOKKPT 60
   |||||
Db 1 MATSLFFMSTQNSVGNPNLLNRLVNVSSGEIRTEITLKSGRKPGSKTGQOKKPT 60

QY 61 LRGMGVAKLQRORIEBEKKOLAAATVGDTSVASISNNATRLPVDPGVWLGFPSSLG 120
   |||||
Db 61 LRGMGVAKLQRORIEBEKKOLAAATVGDTSVASISNNATRLPVDPGVWLGFPSSLG 120

QY 121 SNRIYCGVGSGVMIDPVISPWGFVETSTTHLSSISNPOMFNASSNNRCDTCFKKR 180
   |||||
Db 121 SNRIYCGVGSGVMIDPVISPWGFVETSTTHLSSISNPOMFNASSNNRCDTCFKKR 180

QY 181 LDGDQNNVRSNGGFSKYTIMPPMNGYDQYLLQSDHHQSQGLYDHRHIAAASVSAS 240
   |||||
Db 181 LDGDQNNVRSNGGFSKYTIMPPMNGYDQYLLQSDHHQSQGLYDHRHIAAASVSAS 240

QY 241 STTINPYFNEATNHTGPMEEFGSYMEGNPRNGSGGVKEYEFPFGKYGERVSVVATSSLV 300
   |||||
Db 241 STTINPYFNEATNHTGPMEEFGSYMEGNPRNGSGGVKEYEFPFGKYGERVSVVATSSLV 300

QY 301 GDCSPNTIDLSLKL 314
   |||||
Db 301 GDCSPNTIDLSLKL 314

RESULT 2
081836 PRELIMINARY; PRT: 314 AA.
ID 081836
AC 081836;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE Hypothetical 34.1 kDa protein (NOZZLE/SPOROCTYLESS).
GN M4122.140 OR A14627330.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidops.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Reichert B.J., Barel E., Hoheisel J., Mewes H.W., Mayer K.F.X.,
RA Schueller C., Bevan M.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Reichert B.J., Barel E., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL030978; CAA19727.1;
DR EMBL: AL161566; CAB79588.1;
KW Hypothetical protein.
SQ SEQUENCE 314 AA; 34078 MW; B0236D910544CEC3 CRC64;

Query Match 99.6%; Score 1628; DB 10; Length 314;
Best Local Similarity 99.7%; Pred. No. 3.5e-130;
Matches 313; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MATSLFFMSTQNSVGNPNLLNRLVNVSSGEIRTEITLKSGRKPGSKTGQOKKPT 60
   |||||
Db 1 MATSLFFMSTQNSVGNPNLLNRLVNVSSGEIRTEITLKSGRKPGSKTGQOKKPT 60

QY 61 LRGMGVAKLQRORIEBEKKOLAAATVGDTSVASISNNATRLPVDPGVWLGFPSSLG 120
   |||||
Db 61 LRGMGVAKLQRORIEBEKKOLAAATVGDTSVASISNNATRLPVDPGVWLGFPSSLG 120

QY 121 SNRIYCGVGSGVMIDPVISPWGFVETSTTHLSSISNPOMFNASSNNRCDTCFKKR 180
   |||||
Db 121 SNRIYCGVGSGVMIDPVISPWGFVETSTTHLSSISNPOMFNASSNNRCDTCFKKR 180
```

```
Db 121 SNRIYCGVGSGVMIDPVISPWGFVETSTTHLSSISNPOMFNASSNNRCDTCFKKR 180

QY 181 LDGDQNNVRSNGGFSKYTIMPPMNGYDQYLLQSDHHQSQGLYDHRHIAAASVSAS 240
   |||||
Db 181 LDGDQNNVRSNGGFSKYTIMPPMNGYDQYLLQSDHHQSQGLYDHRHIAAASVSAS 240

QY 241 STTINPYFNEATNHTGPMEEFGSYMEGNPRNGSGGVKEYEFPFGKYGERVSVVATSSLV 300
   |||||
Db 241 STTINPYFNEATNHTGPMEEFGSYMEGNPRNGSGGVKEYEFPFGKYGERVSVVATSSLV 300

QY 301 GDCSPNTIDLSLKL 314
   |||||
Db 301 GDCSPNTIDLSLKL 314

RESULT 3
09LWFL PRELIMINARY; PRT: 320 AA.
ID 09LWFL
AC 09LWFL;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Hypothetical protein (OSJNBa0038J17.1 protein).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0031E09."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
RT clone:OSJNBa0038J17."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP002092; BAA96171.1;
DR EMBL: AP003104; BAB55710.1;
KW Hypothetical protein.
SQ SEQUENCE 320 AA; 34068 MW; AC256F3C99245D2B CRC64;

Query Match 8.0%; Score 131; DB 10; Length 320;
Best Local Similarity 21.1%; Pred. No. 0.0049;
Matches 56; Conservative 33; Mismatches 78; Indels 98; Gaps 9;

QY 53 QOKQKPTLRGMGVAKLERQRTTEEEKKOLAAATVGDTSVASISNNATRLPVDPGVW 112
   |||||
Db 37 KNRKPKVPQRLGVALEKLRTEQKMEGGAALSHGALGGDGLVHLPPPPPSLAL 96

QY 113 Q-----GFPSSLGSNRIYCGVGSGVMIDPVISPW-----GFVE 147
   |||||
Db 97 SALQSRSAADGGFPAALWS-----PVDPIKHPYKRSICPPPPSPMSTGLSL 143

QY 148 TSSTTHLSSISNPOMFNASSNNRCDTCFKKRLDGDQNNVRSNGGFSKYTIMPPMN 207
   |||||
Db 144 TSARHPTEPPSN-OMYSSSGSRSSAAAEER-----GMAAFRTTTSRAPL- 190

QY 208 GYDQYLLQSDHHQSQGLYDHRHIAAASVSASSTT-----INP-----Y 247
   |||||
Db 191 -----VAPFSARTTTTGTGELAGADVDPDLSRYEFRTATN 224

QY 248 FNEATNHTGPMEEFGSYMEGNPRNG 272
   |||||
Db 225 FSNANYSWTSDF-ARCKSKENG 248
```

```

08887
D C Q08887 PRELIMINARY; PRT; 554 AA.
C Q0887;
T T 01-NOV-1996 (TREMBLrel. 01, Created)
T T 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
T T 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
E Chromosome XV reading frame ORF YOR372C.
N NDD1 OR YOR372C.
S Saccharomyces cerevisiae (Baker's yeast).
S Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
S Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
X NCBI_TaxID=4932;
N [1]
P SEQUENCE FROM N.A.
A Deilus H., Hebling U., Hofmann B.;
L Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
N [2]
P SEQUENCE FROM N.A.
A MIPS;
L Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
R EMBL: 275280; CAA99703.1; -.
R SGD: S0005899; NDD1.
Q SEQUENCE 554 AA; 60522 MW; 1CF99115F2B1F23D CRC64;

Query Match      6.7%; Score 109; DB 3; Length 554;
Best Local Similarity 22.7%; Pred. No. 0.78;
Matches 63; Conservative 37; Mismatches 79; Indels 98; Gaps 13;

Y Y 7 FMSTD-QNSVGPNDDLNRTRLVANSSGSEIRTELKSRKGPGSKTGOKKKPTLRGMG 65
    +|||+|||||+:|+:|+:|+:|+:|+:|+:|+:|+:|+:|+:|+:|+:|+:|+:|+:|
b b 71 FANTDAHNSNESSIVENSIL-----PHHQIQOQQOQ----- 103

Y Y 66 VAKLERQRREEKKQLA-----AAVGDTSVASISNNATRLPVDPVGVVLGGFPSSLG 120
    :||:|||||+|||+|||+|||+|||+|||+|||+|||+|||+|||+|||+|||
b b 104 ---QQOQQOQQOQALGLSLVPVAVTDTSETLDIN-----VQPSVVLO-FGNSLP 150

Y Y 121 SNRIYCGVSGGVMD-----PWISPFGVETSTTHELSSINP-OMFNAS 167
    -:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
b b 151 SEFLVASPEQFKEFLLDSPSTNFNFHKTPAKTPLEFV-TDSNGAQQTENPGQQNVF 209

Y Y 168 SNNRCDTCFKKELDGQNNVRNG-----GFESKYTWIPPMNGYQYLLOSDDH 219
    |||||+|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
b b 210 SN-----VDLANLKSNKTPSSSCTGAFSK-----TPLSKIDMNL----- 245

Y Y 220 QRSGGLFYDHRIAARASVSASSSTNPYPNEATNHGT 256
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
b b 246 -----MENOPTSPSKRFSSLSLTPTYGRKIINDVG 276

RESULT 5,
9WJX4 PRELIMINARY; PRT; 575 AA.
C Q9VJX4
C Q9VJX4;
T T 01-MAY-2000 (TREMBLrel. 13, Created)
T T 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
T T 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
E CG16851 protein.
N CG16851.
S Drosophila melanogaster (Fruit fly).
S Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
S Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
S Ephydroidea; Drosophilidae; Drosophila.
X NCBI_TaxID=7227;
N [1]
P SEQUENCE FROM N.A.
A STRAIN=BERKELEY;
C MEDLINE=19196006; PubMed=10731132;
X X Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
A Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
A George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
A Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
A Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
A Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

```

```

RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berhan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J.J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA de Fabios B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Keenison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Lilang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulyov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclic J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Renington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RA Science 287:2185-2195(2000).
RA EMBL; AF003640; AAF5311.1; -.
RA FlyBase; FBgn0032528; CG16851.
RA SEQUENCE 575 AA; 61004 MW; 3EB0045E316A0B07 CRC64;
Query Match 6.4%; Score 104; DB 5; Length 575;
Best Local Similarity 23.5%; Pred. No. 2.2;
Matches 55; Conservative 31; Mismatches 86; Indels 62; Gaps
QY 43 RGRKPGSKTQXQKRPRLRGMGVAKLERIBEEKQLAAATVCDTSSVASISNNATRL 102
DB 27 RGRRCVSQDQOQQQATCHRHW-----VKNRKIQD-----ASUGGSDDEDFTLG 69
QY 103 PVPVDPGVVLQGFPSLGSNRIYCGVGSGQVMIDPVTSPWGFVETSTHTLSSISNPQ 162
DB 70 -----VLRG-PSIFANAFLY---VGLGTVALGLVIAFGVTGKCFKTELRLIGPSL 117
QY 163 MFWASSNNRCDTCFKKKR-----LDGQDNVVRNSGGFSKYTMIPPPMNGYQYLLQS 216
DB 118 IGHKIANGKQQLIKKTRKKNKIDADHTTSLRLNESKRPYPIA----- 163
QY 217 DHQRSGFLYDHRHRAAASVSASS-----TTIN---PYFNEATNHTGPME 259
DB 164 --HKRSQKMLNMGMEALRQIATTSLEFMQNEQKTAIRKRVVPIINEPESGDPAPLE 215
RESULT 6
Q9EPU2 PRELIMINARY; PRT; . 1082 AA.
ID AC Q9EPU2
IC Q9EPU2
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Nucleur receptor coactivator AIB1 (fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
[1]
SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RA Birsby R.M., Long X., Nephew K.P.;

```

RT "PD98059 induces estrogen receptor-coactivator interaction and
 RT reporter gene transcription."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR ENBL; AF322224; AAG42837.1; -
 KW Receptor.
 FT NON_TER 1
 SQ SEQUENCE 1082 AA; 115131 MW; BA35D946CBC8C080 CRC64;

Query Match 6.3%; Score 103.5; DB 11; Length 1082;
 Best Local Similarity 21.7%; Pred. No. 5.9;
 Matches 73; Conservative 34; Mismatches 150; Indels 79; Gaps 10;

QY 4 SLFPMSTQNSVGNPNLLNRLNRLVNVSSGEIRTEITLTKSRGKPGSKTGQOKKPTLRG 63
 Db 70 STIFLQRENGC-RNPIL-----QDKIRPPAAG 98
 QY 64 MGAKLERQRIE-----EEKQLAAATVGTGTSVASISNNATRLPVPVDPGVV 111
 Db 99 CGMSLSPSQVQLGSRITYGVADPSNTGOMAGARYGASSVASLT-----PGQS 147
 QY 112 LQGFPSLSGNRYC-----GGVSGQVMDPVISPMPGVFVETSTTHLSSISNPQMF 164
 Db 148 LQS-PSSYQNSYGLNMSSPHGSPGLGNQNMISPRNRGSPKMAHQFSPAAGVHSP 206
 QY 165 NASSNNRCDTCFKKRLDGDQNNVRSNGGGFSKYTWIDPP-----MNGYDQYLQ 215
 Db 207 MGSNGTNGSHSFSSLSALQ---AISEGVGTSLLSLSPGPKLDNSPNMN-INQPSKA 262
 QY 216 SDHQRSQGLYDHRITARAASVSASTTINPYNEATNHTGPMEEFGSYMEGNPRNGSG 275
 Db 263 SSODSKSPGLYCEQNPVSSVCPNSRDPPTVTKENKNSGEASE-----TPROPLESKH 318
 QY 276 VKREYFFPGYGRVSVVATTSVLGDCSPNTIDL 311
 Db 319 KKLQLLTSSDDRGHSLTNSPLDSCNCKDSSISVT 354

RESULT 7
 Q9SBF4 PRELIMINARY; PRT; 276 AA.

ID Q9SBF4
 AC Q9SBF4
 DT 01-MAY-2000 (TREMREL. 13, Created)
 DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
 DT 01-MAR-2002 (TREMREL. 20, Last annotation update)
 DE Putative transcription factor (Fragment).
 GN MYB86.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsia.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=99056348; PubMed=9839469;
 RA Kranz H.D., Denekamp M., Greco R., Jin H., Leyva A., Meissner R.C.,
 RA Petroni K., Urzaizqui A., Bevan M., Martin C., Smeekens S.,
 RA Tonelli C., Paz-Ares J., Weisshaar B.;
 RT "Towards functional characterisation of the members of the R2R3-MYB
 RL gene family from Arabidopsis thaliana.";
 RL Plant J. 16:263-276(1998).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
 DR EMBL; AF062913; AAC83635.1; -
 DR HSP; Q03237; IASU.
 DR InterPro; IPR001005; Myb_DNA_binding.
 DR Pfam; PF00249; myb_DNA_binding; 1.
 DR SMART; SM00395; SANT; 1.
 DR PROSITE; PS00334; MYB_2; 1.
 DR PROSITE; PSS0090; MYB_3; 1.
 KW DNA-binding; Nuclear protein.
 FT NON_TER 1
 SQ SEQUENCE 276 AA; 30888 MW; FA22CDAF63C53953 CRC64

Query Match 6.3%; Score 103; DB 10; Length 276;
 Best Local Similarity 23.3%; Pred. No. 0.95;
 Matches 60; Conservative 41; Mismatches 107; Indels 50; Gaps 12;

QY 25 TRLVNSSGEIR---TETLKSRRKPGSKTGQOKKPTLRGMGVAKLERQRIEKKOL 81
 Db 18 TRLPGRTDEIKFNWNSCLKKLRKGI---DPTHKPLITN-----ELQSLNYDOKL 68
 QY 82 AAATVGTGTSVASISNNATRLPVPVDPGVV---QG---SSLGNSRIYCGVSGG 132
 Db 69 TSSEV--VKSTGGINLH-----DQSMVSSQQGPMFWPANTTTTNGNSAFSSNT 119
 QY 133 QVMIDPVISPGFVETSTTHLSSISNPQMFNASSNRCDCFCFKKRLDGDQNNVRSN 192
 Db 120 TTVSDQIVSLISSMTSSSTPTMTSNFSPAPNNWEOLNYCNTV-----PSQNSIYSA 172
 QY 193 GGFSSKYTWIPPMNGYDQYLQSDHHSQGLYDHRITARAASVSASTTI-----244
 Db 173 FRG-NOYTEASQTMNNNPLVDQHHHQDKMSWASE--ILHYTERSQSSETVIEAEVDP 229
 QY 245 --NPYFNEATNHTGPMEE 260
 Db 230 IANYWRSASSSSPNQE 247

RESULT 8
 Q9VXV8 PRELIMINARY; PRT; 1693 AA.

ID Q9VXV8
 AC Q9VXV8
 DT 01-MAY-2000 (TREMREL. 13, Created)
 DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
 DT 01-MAR-2002 (TREMREL. 20, Last annotation update)
 DE CG9176 protein.
 GN CNG1 OR CG9176.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simonsen M., Skupski M.P., Smith T.,
 RA Stange C., Strydom A., Tabor C., Tabor J., Tabor M., Tabor R.,

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

M nucleic - nucleic search, using sw model

un on: January 25, 2003, 21:04:51 : Search time 262 seconds
(without alignments)
8114.070 Million cell updates/sec

file: US-09-701-023-1_COPY_81_1024
erfect score: 944
equence: 1 tggcgaactctctctcttc.....attgtccttggaagctttaa 944

coring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

eached: 2185239 seqs, 1125999159 residues

otal number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : N_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

ALIGNMENTS

RESULT 1
AAA94065
ID AAA94065 standard; cDNA; 1302 BP.
XX
AC AAA94065;
XX
DT 30-JAN-2001 (first entry)
XX
DE Arabidopsis thaliana sporocyteless spl coding sequence.
XX
KW Sporocyteless; spl; meiocyte formation; plant sterility; seedless fruit;
KW pollenless flower; Ds element; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 80..1024
FT /tag= a
FT /product= "SPL"
FT /trans_except= (pos:146..148,aa:Irg)
XX
PN WO200056907-A1.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-1999; 99WO-SG00023.
XX
PR 22-MAR-1999; 99WO-SG00023.
XX
PA (MOLE-) INST MOLECULAR AGROBIOLOGY.
XX
PI Ye D, Yang W, Sundaresan V, Xu J;

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	944	100.0	1302	21 AAA94065	Arabidopsis thalia
2	732.8	77.6	4071	21 AAA94067	Arabidopsis thalia
3	127.4	13.5	271	21 AAA94066	Arabidopsis thalia
C 4	37.2	3.9	4590	22 AAH24065	Yeast AOD9604-asso
C 5	36.2	3.8	9780	22 AAK72360	Human immune/haema
C 6	35.8	3.8	465237	24 ABQ87681	Human oestrogen re
C 7	35.8	3.8	465237	24 ABA90193	Human oestrogen re
C 8	34.6	3.7	583	24 ABN64389	Human cancer relat
9	34.6	3.7	4604	24 ABQ70984	Listeria monocytog

```
XX DR WPI; 2000-594578/56.
XX DR P-PSDB; AAB36111.
XX PT Sporocytless nucleic acids and polypeptides, useful for controlling
XX PT sporocyte or meiocyte formation in plants for producing transgenics
XX PT that bear seedless fruits and/or pollenless flowers -
XX PS
XX PS Claim 2; Fig 2; 65pp; English.
XX CC The present sequence is the coding sequence for the Arabidopsis thaliana
XX CC sporocytless (SPL) protein. This protein is involved in the formation of
XX CC microsporocytes in male plants and megasporocytes in female plants. The
XX CC mutant gene was isolated from a collection of transposants due to its
XX CC male and female sterile phenotype. The mutation is caused by the
XX CC insertion of the Ds element between bases 411 and 412 of the gene,
XX CC causing a 4 base pair duplication. The identification of this element
XX CC phenotype in plants caused by the insertion of this element means that it
XX CC can be used to produce transgenic plants which contain the Ds element or
XX CC antisense sequences to the gene, enabling the production of seedless
XX CC fruits, pollenless flowers and plants with a larger biomass.
XX SQ Sequence 1302 BP; 402 A; 273 C; 275 G; 352 T; 0 other;

Query Match 100.0%; Score 944; DB 21; Length 1302;
Best Local Similarity 100.0%; Pred. No. 2.2e-299;
Matches 944; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGACTCTCTCTTCTTCATGTCACAGATCAAAATCCGTCGGAAACCCAAACGATC 60
DB 81 TGGCGACTCTCTCTTCTTCATGTCACAGATCAAAATCCGTCGGAAACCCAAACGATC 140
QY 61 TTCGAGAAACACCCGCTTGTGCTCAATAGCTCCGCGGAGATCCGGACAGACACTGA 120
DB 141 TTCGAGAAACACCCGCTTGTGCTCAATAGCTCCGCGGAGATCCGGACAGACACTGA 200
QY 121 AGAGTCGTGTCGGAACACAGGATCGAGACAGGTCAGCAAAACAGAAACACCAACG 180
DB 201 AGAGTCGTGTCGGAACACAGGATCGAGACAGGTCAGCAAAACAGAAACACCAACG 260
QY 181 TGAGAGGAATGGGTGTAGCAAGCTCGAGCTCAGAGATCGAAGAAAGAAAGCAAC 240
DB 261 TGAGAGGAATGGGTGTAGCAAGCTCGAGCTCAGAGATCGAAGAAAGAAAGCAAC 320
QY 241 TCGCGCGCGCCACAGTCGGGAGACACGTCATCAGTAGCATCGATCTCTAAACACGCTACCC 300
DB 321 TCGCGCGCGCCACAGTCGGGAGACACGTCATCAGTAGCATCGATCTCTAAACACGCTACCC 380
QY 301 GTTTACCCGTACCGGTAGACCCGGGTGTGTGCTACAGGCTCCCAAGCTCACTCGGA 360
DB 381 GTTTACCCGTACCGGTAGACCCGGGTGTGTGCTACAGGCTCCCAAGCTCACTCGGA 440
QY 361 GCACAGGATCTATTGTGGTGGAGTCGGGTGCGGTATGATCGACCCGGTTATT 420
DB 441 GCACAGGATCTATTGTGGTGGAGTCGGGTGCGGTATGATCGACCCGGTTATT 500
QY 421 CTCATGGGGTTTGTGTGAGACCTCTCCACTACTCATGAGCTCTCTCAATCTCAATC 480
DB 501 CTCATGGGGTTTGTGTGAGACCTCTCCACTACTCATGAGCTCTCTCAATCTCAATC 560
QY 481 CTCAAATGTTTACGGTCTCTTCCAAATAGCGTGTGACACTTGTCTTCAAGAAACGTT 540
DB 561 CTCAAATGTTTACGGTCTCTTCCAAATAGCGTGTGACACTTGTCTTCAAGAAACGTT 620
QY 541 TGGATGTCATCAGATAATGTAGTTCGATCCAAACGGTGGTGGATTTTCAATACACAA 600
DB 621 TGGATGTCATCAGATAATGTAGTTCGATCCAAACGGTGGTGGATTTTCAATACACAA 680
QY 601 TGATTCCTCTCTCCGATGAACGGTACGATCATCTTCTTCAATCAGATCATCATCAGA 660
DB 681 TGATTCCTCTCTCCGATGAACGGTACGATCATCTTCTTCAATCAGATCATCATCAGA 740
QY 661 GGAGCAAGGTTTCTCTTATGATCATAGAAATCCGTAGACAGCTTCAGTTCTTCTGCTTCA 720
```

3 larger biomass.
x
2 Sequence 4071 BP; 1306 A; 742 C; 692 G; 1331 T; 0 other;
Query Match 77.6%; Score 732.8; DB 21; Length 4071;
Best Local Similarity 83.2%; Pred. No. 1.1e-229;
Matches 942; Conservative 0; Mismatches 2; Indels 188; Gaps 2;
1 TCGCGACTTCTCTCTTCTCATGTCAACAGATCAAAACTCCGTCGGAACCCAAACGATC 60
2 2691 TGGCGACTTCTCTCTTCTCATGTCAACAGATCAAAACTCCGTCGGAACCCAAACGATC 2750
3
4 61 TTCTGAGAAACACCGTCTTGTGCTCAATAGTCCGGGAGATCCGGACAGACACTGA 120
5 2751 TTCTGAGAAACACCGTCTTGTGCTCAACAGTCCGGGAGATCCGGACAGACACTGA 2810
6
7 121 AGAGTCGTGGTCGGAACACGAGATCGAGACAGGTTCAGCAAAACAGAGAACAACGAT 180
8 2811 AGAGTCGTGGTCGGAACACGAGATCGAGACAGGTTCAGCAAAACAGAGAACAACGAT 2870
9
10 181 TGAGAGGAATGGTGTAGCAAAAGCTCGAGCGTCAGAGATCGAAGAAAGAAAGCAAC 240
11 2871 TGAGAGGAATGGTGTAGCAAAAGCTCGAGCGTCAGAGATCGAAGAAAGAAAGCAAC 2930
12
13 241 TCGCCGCCGCCACAGTCGGAGACAGCTCATCAGTAGCATCGATCTTAACACGCTACCC 300
14 2931 TCGCCGCCGCCACAGTCGGAGACAGCTCATCAGTAGCATCGATCTTAACACGCTACCC 2990
15
16 301 GTTATCCCGTACCGTAGACCGGGTGTGCTCAACAGGTTCCTCAAGCTCACTCGGGA 360
17 2991 GTTATCCCGTACCGTAGACCGGGTGTGCTCAACAGGTTCCTCAAGCTCACTCGGGA 3050
18
19 361 GCAACAGGATCATTTGTGTGAGTCGGGTCCGGTCAGGTATGATCGACCCGGTATTT 420
20 3051 GCAACAGGATCATTTGTGTGAGTCGGGTCCGGTCAGGTATGATCGACCCGGTATTT 3110
21
22 421 CTCATGGGGTTTGTGAGACCTCTCCACATCATCATGAGCTCTTCAATCTCAATC 480
23 3111 CTCATGGGGTTTGTGAGACCTCTCCACATCATCATGAGCTCTTCAATCTCAATC 3170
24
25 481 CTCAAATGTTTAAACGTTCTTCCATAATCGCTGTGACACTGCTTCA----- 528
26 3171 CTCAAATGTTTAAACGTTCTTCCATAATCGCTGTGACACTGCTTCAAGGTTGTTGT 3230
27 529 ----- 528
28 3231 TTTTAAATCGTTTTCATCAACATGATTGATATATATATATAGTTTTCGACTTGAAAAGTT 3290
29 529 -----AGAGAAACGGTTTGGATGTCATCAGATAATG 561
30 3291 TTGATTTTATTTATGTAARAACCTCAGAGAAGCGTTTGGATGGTATCAGATAATG 3350
31 562 TAGTTCGATCCAACGGTGGTGGATTTTCAAAATACAAATGATTCCTCTCCGATGAAG 621
32 3351 TAGTTCGATCCAACGGTGGTGGATTTTCAAAATACAAATGATTCCTCTCCGATGAAG 3410
33 622 GCTACGATCAGTATCTTCTCAATCAGATCATCATCAGAGGACCAAGTTTCCCTTATG 681
34 3411 GCTACGATCAGTATCTTCTCAATCAGATCATCATCAGAGGACCAAGTTTCCCTTATG 3470
35 682 ATCATAGATCCGTAGACGAGTTCAGTTCTGCTCTAGTACTACTATTAATCTCTTAT 741
36 3471 ATCATAGATCCGTAGACGAGTTCAGTTCTGCTCTAGTACTACTATTAATCTTAT 3530
37 742 TCAACGAGGCAACAATATCATC----- 763
38 3531 TCAACGAGGCAACAATATCATCAGGTACTAAGTAGTAGTCCATTTATTAATCATATATA 3590
39 764 -----GGGACCAAT 772
40 3591 GGTATATATGATATAAAGTGTGATCTTATTTGATTAACTGGTGGTTTAGGGACCAAT 3650
41 773 CGAGGATTTGGAGCTACATGGAAGAAACCTTAGAATGATCAGGAGTGTGAGGA 832

Db 3651 GGAGGAATTTGGAGCTACATGGGAAGAAACCTTAGAATGGATCAGGAGTGTGAAGGA 3710
QY 833 GTACGAGTTTTTTCGGGGAAATATGTTCAAGAGATTTTCAGTGGGTACAACTCGTC 892
Db 3711 GTACGAGTTTTTTCGGGGAAATATGTTCAAGAGATTTTCAGTGGGTAAACCTCGTC 3770
QY 893 ACTCGTAGGTGATTCAGCTCCTCAATACCATTTGATTGCTTGAAGCTTTAA 944
Db 3771 ACTCGTAGGTGATTCAGCTCCTCAATACCATTTGATTGCTTGAAGCTTTAA 3822
RESULT 3
AAA94066
ID AAA94066 standard; DNA; 271 BP.
XX
XX AAA94066;
XX
XX 30-JAN-2001 (first entry)
XX
DE Arabidopsis thaliana sporocyteless spl Ds element sequence.
XX
KW Sporocyteless; spl; meocyte formation; plant sterility; seedless fruit;
KW pollenless flower; Ds element; ds.
XX
OS Arabidopsis thaliana.
XX
XX Key Location/Qualifiers
FH misc_feature 60..63 /*tag= a
FT /*note= "4 base pair duplication due to Ds element
FT insertion" 134..137
FT misc_feature /*tag= b
FT /*note= "4 base pair duplication due to Ds element
FT insertion" 64..133
FT misc_feature /*tag= c
FT /*note= "Ds element"
XX WO200056907-A1.
XX
XX 28-SEP-2000.
XX
XX 22-MAR-1999; 99WO-SG00023.
XX
XX 22-MAR-1999; 99WO-SG00023.
XX (MOLE-) INST MOLECULAR AGROBIOLOGY.
XX
XX Ye D, Yang W, Sundaresan V, Xu J;
XX WPI; 2000-594578/56.
XX
XX Sporocyteless nucleic acids and polypeptides, useful for controlling
XX sporocyte or meocyte formation in plants for producing transgenics
XX that bear seedless fruits and/or pollenless flowers -
XX
XX Claim 4; Fig 1A; 65pp; English.
XX
XX The present sequence comprises part of the coding sequence for the
XX Arabidopsis thaliana sporocyteless (SPL) protein containing the Ds
XX element. This protein is involved in the formation of microsporocytes in
XX male plants and megasporocytes in female plants. The mutant gene
XX (containing the element) was isolated from a collection of transposants
XX due to its male and female sterile phenotype. The mutation is caused by
XX the insertion of the Ds element between bases 411 and 412 of the gene,
XX causing a 4 base pair duplication. The identification of the sterile
XX phenotype in plants caused by the insertion of this element means that it
XX can be used to produce transgenic plants which contain the Ds element or
XX antisense sequences to the gene, enabling the production of seedless
XX fruits, pollenless flowers and plants with a larger biomass.

556 ATAATGCTAGTTCGATCCAACTGGTGGATTTTCGAAATACACAATGATTCCTCCTCGGA 615
223 TBSRTBTTSRBSRTSRHSRTTBTBTBTTHVANTBGTSRAGDISHINTYNASTA 164
616 TGAACGGCTACGATCAGTATCTTCCTCAATCAGATCAATCAGAGAGCCCAAGGTTTCC 675
163 TATDASBYDADAWYDNGHGTGNTKCHVTDGNTTSHSYCNASSNGYGTSTYYDNNSTYSR 104
676 TTTATGATCATAGATCGCTAGACGAGCTT 705
103 VHAASVSRAKGDTRAKGDTDSYRAKGDTY 74

RESULT 5

AK72360/c
AAK72360 standard; DNA; 9780 BP.

AAK72360;

06-NOV-2001 (first entry)

Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27172.
Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
cytostatic; gene therapy; vaccine; metastasis; ds.

Homo sapiens.

WO200157182-A2.

09-AUG-2001.

17-JAN-2001; 2001WO-US01354.

31-JAN-2000; 2000US-0179065.

04-FEB-2000; 2000US-0180628.

24-FEB-2000; 2000US-0184664.

02-MAR-2000; 2000US-0186350.

16-MAR-2000; 2000US-0189874.

17-MAR-2000; 2000US-0190076.

18-APR-2000; 2000US-0198123.

19-MAY-2000; 2000US-0205515.

07-JUN-2000; 2000US-0209467.

28-JUN-2000; 2000US-0214886.

30-JUN-2000; 2000US-0215135.

07-JUL-2000; 2000US-0216647.

07-JUL-2000; 2000US-0216880.

11-JUL-2000; 2000US-0217487.

11-JUL-2000; 2000US-0217496.

14-JUL-2000; 2000US-0218290.

26-JUL-2000; 2000US-0220963.

26-JUL-2000; 2000US-0220964.

14-AUG-2000; 2000US-0224518.

14-AUG-2000; 2000US-0224519.

14-AUG-2000; 2000US-0225213.

14-AUG-2000; 2000US-0225214.

14-AUG-2000; 2000US-0225266.

14-AUG-2000; 2000US-0225267.

14-AUG-2000; 2000US-0225268.

14-AUG-2000; 2000US-0225270.

14-AUG-2000; 2000US-0225447.

14-AUG-2000; 2000US-0225757.

14-AUG-2000; 2000US-0225758.

14-AUG-2000; 2000US-0225759.

18-AUG-2000; 2000US-0226279.

22-AUG-2000; 2000US-0226681.

22-AUG-2000; 2000US-0226688.

22-AUG-2000; 2000US-0227182.

23-AUG-2000; 2000US-0227009.

30-AUG-2000; 2000US-0228924.

01-SEP-2000; 2000US-0229287.

01-SEP-2000; 2000US-0229343.

01-SEP-2000; 2000US-0229344.
01-SEP-2000; 2000US-0229345.
05-SEP-2000; 2000US-0229509.
05-SEP-2000; 2000US-0229513.
06-SEP-2000; 2000US-0230437.
06-SEP-2000; 2000US-0230438.
08-SEP-2000; 2000US-0231242.
08-SEP-2000; 2000US-0231243.
08-SEP-2000; 2000US-0231244.
08-SEP-2000; 2000US-0231413.
08-SEP-2000; 2000US-0231414.
08-SEP-2000; 2000US-0232080.
08-SEP-2000; 2000US-0232081.
12-SEP-2000; 2000US-0231968.
14-SEP-2000; 2000US-0232397.
14-SEP-2000; 2000US-0232398.
14-SEP-2000; 2000US-0232399.
14-SEP-2000; 2000US-0232400.
14-SEP-2000; 2000US-0232401.
14-SEP-2000; 2000US-0233063.
14-SEP-2000; 2000US-0233064.
14-SEP-2000; 2000US-0233065.
21-SEP-2000; 2000US-0234223.
21-SEP-2000; 2000US-0234274.
25-SEP-2000; 2000US-0234997.
25-SEP-2000; 2000US-0234998.
26-SEP-2000; 2000US-0235484.
27-SEP-2000; 2000US-0235834.
27-SEP-2000; 2000US-0235836.
29-SEP-2000; 2000US-0236327.
29-SEP-2000; 2000US-0236367.
29-SEP-2000; 2000US-0236368.
29-SEP-2000; 2000US-0236369.
29-SEP-2000; 2000US-0236370.
02-OCT-2000; 2000US-0236802.
02-OCT-2000; 2000US-0237037.
02-OCT-2000; 2000US-0237038.
02-OCT-2000; 2000US-0237039.
02-OCT-2000; 2000US-0237040.
13-OCT-2000; 2000US-0239935.
13-OCT-2000; 2000US-0239937.
20-OCT-2000; 2000US-0240960.
20-OCT-2000; 2000US-0241221.
20-OCT-2000; 2000US-0241785.
20-OCT-2000; 2000US-0241786.
20-OCT-2000; 2000US-0241787.
20-OCT-2000; 2000US-0241808.
20-OCT-2000; 2000US-0241809.
20-OCT-2000; 2000US-0241826.
01-NOV-2000; 2000US-0244617.
08-NOV-2000; 2000US-0246474.
08-NOV-2000; 2000US-0246475.
08-NOV-2000; 2000US-0246476.
08-NOV-2000; 2000US-0246477.
08-NOV-2000; 2000US-0246478.
08-NOV-2000; 2000US-0246523.
08-NOV-2000; 2000US-0246524.
08-NOV-2000; 2000US-0246525.
08-NOV-2000; 2000US-0246526.
08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246609.
08-NOV-2000; 2000US-0246610.
08-NOV-2000; 2000US-0246611.
08-NOV-2000; 2000US-0246613.
17-NOV-2000; 2000US-0249207.
17-NOV-2000; 2000US-0249208.
17-NOV-2000; 2000US-0249209.
17-NOV-2000; 2000US-0249210.
17-NOV-2000; 2000US-0249211.
17-NOV-2000; 2000US-0249212.
17-NOV-2000; 2000US-0249213.

PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-483426/52.
DR
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX Disclosure: SEQ ID NO 27172; 3071pp + Sequence Listing: English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients' own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 9780 BP; 2899 A; 1706 C; 1763 G; 3412 T; 0 other;

Query Match 3.8%; Score 36.2; DB 22; Length 9780;
Best Local Similarity 51.6%; Pred. NO. 1.7;
Matches 83; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
OY 626 CGATCAGTATCTTTCATCATCATCATCATCAGAGGCCAAGCTTCCCTTTATGATCA 685
Db 1124 CCACTATTATTACCTTAGTTGTTAATAAATAAGATATATGGAATTTCCAAAATAATAA 1065
OY 686 TAGAATCCGTCAGGAGCGTTTCAGTTTCCTTCTAGTACTACTATTAATCTTATTTCAA 745
Db 1064 AAAAGATATTACTGCAAAATAGCTTACTATCTACCTTACTACTATTTTATTTAA 1005
OY 746 CGAGGCACAAATATGACGGGACCAATGAGGAATTTGGGA 786

RESULT 6

ABQ87681/c
ID ABQ87681 standard; DNA; 465237 BP.
XX
AC ABQ87681;
XX
DT 18-SEP-2002 (first entry)
XX
DE Human oestrogen receptor alpha gene.
XX
KW Human; oestrogen; receptor; oestrogen receptor alpha; cytostatic;
KW osteopathic; cardiant; cancer; osteoporosis; cardiovascular disorder;
KW chromosome 6q25.1; gene; ds.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
FT variation replace(18783,T)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(18937,C)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT exon 18941..19032
FT /tag= b
FT /number= 1
FT intron 19033..52817
FT /tag= b
FT /number= 1
FT variation /cons_splice= (5'site:YES,3'site:NO)
FT replace(19034,C)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT exon 52818..52940
FT /tag= b
FT /number= 2
FT variation replace(52877,A)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(52901,A)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT intron 52941..64149
FT /tag= b
FT /number= 2
FT /cons_splice= (5'site:YES,3'site:NO)
FT exon 64150..64280
FT /tag= b
FT /number= 3
FT intron 64281..166227
FT /tag= b
FT /number= 3
FT variation /cons_splice= (5'site:YES,3'site:NO)
FT replace(64331,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT exon 166228..166322
FT /tag= b
FT /number= 4
FT intron 166323..168001
FT /tag= b
FT /number= 4
FT variation replace(167950,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(167989,G)
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT exon 168002..168120
FT /tag= b

T	variation	replace(168054,G)
T		/*tag= a
T	intron	/*standard_name= "Single nucleotide polymorphism"
T		168121..169542
T		/*tag= b
T		/number= 5
T		/*cons_splice= (5'site:YES,3'site:NO)
T	exon	169543..169825
T		/*tag= b
T		/number= 6
T	variation	replace(169812,G)
T		/*tag= a
T		/*standard_name= "Single nucleotide polymorphism"
T	variation	replace(169823,G)
T		/*tag= a
T		/*standard_name= "Single nucleotide polymorphism"
T	intron	169826..169866
T		/*tag= b
T		/number= 6
T		/*cons_splice= (5'site:YES,3'site:NO)
T	exon	169867..170678
T		/*tag= b
T		/number= 7
T	variation	replace(170035,A)
T		/*tag= a
T		/*standard_name= "Single nucleotide polymorphism"
T	variation	replace(170068,T)
T		/*tag= a
T		/*standard_name= "Single nucleotide polymorphism"
T	variation	replace(170256,C)
T		/*tag= a
T		/*standard_name= "Single nucleotide polymorphism"
T	variation	replace(170368,G)
T		/*tag= a
T		/*standard_name= "Single nucleotide polymorphism"
T	variation	replace(170487,C)
T		/*tag= a
T		/*standard_name= "Single nucleotide polymorphism"
T	intron	170679..204911
T		/*tag= b
T		/number= 7
T	exon	204912..205102
T		/*tag= b
T		/number= 8
T	intron	205103..242969
T		/*tag= b
T		/number= 8
T	exon	242970..243086
T		/*tag= b
T		/number= 9
T	variation	replace(243055,T)
T		/*tag= a
T		/*standard_name= "Single nucleotide polymorphism"
T	intron	243087..306167
T		/*tag= b
T		/number= 9
T	variation	replace(243187,C)
T		/*tag= a
T		/*standard_name= "Single nucleotide polymorphism"
T	exon	306168..306503
T		/*tag= b
T		/number= 10
T	variation	replace(306292,A)
T		/*tag= a
T		/*standard_name= "Single nucleotide polymorphism"
T	variation	replace(306382,G)
T		/*tag= a
T		/*standard_name= "Single nucleotide polymorphism"
T	intron	306504..373639
T		/*tag= b
T		/number= 10
T	exon	373640..373778
T		/*tag= b

	/number= 11	
FT	373779..422963	
FT	/*tag= b	
FT	/number= 11	
FT	422964..423097	
FT	/*tag= b	
FT	/number= 12	
FT	replace(423067,C)	
FT	/*tag= a	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	423098..456353	
FT	/*tag= b	
FT	/number= 12	
FT	replace(423149,G)	
FT	/*tag= a	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	replace(423163,G)	
FT	/*tag= a	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	replace(423220,A)	
FT	/*tag= a	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	replace(423232,G)	
FT	/*tag= a	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	replace(423258,G)	
FT	/*tag= a	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	456354..456537	
FT	/*tag= b	
FT	/number= 13	
FT	456538..460700	
FT	/*tag= b	
FT	/number= 13	
FT	replace(459706,C)	
FT	/*tag= a	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	replace(459832,A)	
FT	/*tag= a	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	replace(459913,G)	
FT	/*tag= a	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	replace(460024,G)	
FT	/*tag= a	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	replace(460056,T)	
FT	/*tag= a	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	replace(460159,G)	
FT	/*tag= a	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	replace(460553,C)	
FT	/*tag= a	
FT	/standard_name= "Single nucleotide polymorphism"	
FT	replace(460564,A)	
FT	/*tag= a	
FT	/standard_name= "Single nucleotide polymorphism"	
FT		
Query Match	3.8%; Score 35.8; DB 24; Length 465237;	
Best Local Similarity	52.9%; Pred. No.18;	
Matches	99; Conservative 0; Mismatches 87; Indels 1; Gaps 1;	
QY	97 GCGAGATCCGACAGACACTGAAGTAGTCGTGGTCCGAAACCAGAGATCGAAGCAGCTGC 156 	
Ddb	2868 GAGACAGACAGAGAAGAAAGAGAGCGCGGATAGTAAGGAAGGAAGGAAGGAAAGG-A 2810	
QY	157 AGCAAAAACAAGAAACCAACGTTTGAGAGGAATGGGTGTAGCAAAGCTCGACGCGTCAGA 216 	
Ddb	2809 AGCAAGGAGGAGGAAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2750	
QY	217 GAATCGAAGAAAGAAAGCAACTCCGCCGCCACAGTCGGAGACACGTCATCATGASTAG 276 	

C meningococcal septicaemia, cancer, such as breast cancer, ovarian cancer
C or colon cancer. Many other examples of diseases are given in the
C specification. The present sequence encodes human itk.
X
Q Sequence 4221 BP; 1159 A; 913 C; 991 G; 1158 T; 0 other;

Query Match 3.6%; Score 34.2; DB 22; Length 4221;
Best Local Similarity 52.4%; Pred. No. 4.9;
Matches 75; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
Y 720 AGTACTACTATTATCTTATTTCACGAGGCAACAAATCATACGAGGCAATGTGAGGAA 779
b 2080 ATTGCTATTATTTATTCTTTTATTACATGAATCTAAAGTTTATGTTCCAGGACTTTT 2139
Y 780 TTGGAGGTACATGGAAGGAAACCCTAGAAATGGATCAGGAGGTGTGAAGGATACGAG 839
b 2140 TATTGACCAACAACACAGTATCCCGAGGATATGAGGCAAGGGAACAAAGACCATGAG 2199
Y 840 TTTTTCGGGGAAATATGGTGA 862
b 2200 TCTTTTCCAAGAAACTGGTGA 2222

RESULT 13

AK80081
D AK80081 standard; DNA; 11363 BP.
X
C AAK80081;
T
X T 07-NOV-2001 (first entry)
X
E Human Immune/haematopoietic antigen genomic sequence SEQ ID NO:34893.
X
W Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
W cytostatic; gene therapy; vaccine; metastasis; ds.
X Homo sapiens.
X
X WO200157182-A2.
X
X
X
D 09-AUG-2001.
X
F 17-JAN-2001; 2001WO-US01354.
X

31-JAN-2000; 2000US-0179065.

04-FEB-2000; 2000US-0180828.

24-FEB-2000; 2000US-0184564.

02-MAR-2000; 2000US-0186350.

16-MAR-2000; 2000US-0189874.

17-MAR-2000; 2000US-0190076.

18-APR-2000; 2000US-0198123.

19-MAY-2000; 2000US-0205515.

07-JUN-2000; 2000US-0209467.

28-JUN-2000; 2000US-0214886.

30-JUN-2000; 2000US-0215135.

07-JUL-2000; 2000US-0216647.

07-JUL-2000; 2000US-0216880.

11-JUL-2000; 2000US-0217487.

11-JUL-2000; 2000US-0217496.

14-JUL-2000; 2000US-0218290.

26-JUL-2000; 2000US-0220963.

26-JUL-2000; 2000US-0220964.

14-AUG-2000; 2000US-0224518.

14-AUG-2000; 2000US-0224519.

14-AUG-2000; 2000US-0225213.

14-AUG-2000; 2000US-0225214.

14-AUG-2000; 2000US-0225266.

14-AUG-2000; 2000US-0225267.

14-AUG-2000; 2000US-0225268.

14-AUG-2000; 2000US-0225270.

14-AUG-2000; 2000US-0225447.

14-AUG-2000; 2000US-0225757.

14-AUG-2000; 2000US-0225758.

14-AUG-2000; 2000US-0225759.
18-AUG-2000; 2000US-0226279.
22-AUG-2000; 2000US-0226681.
22-AUG-2000; 2000US-0226686.
22-AUG-2000; 2000US-0227182.
23-AUG-2000; 2000US-0227009.
30-AUG-2000; 2000US-0228924.
01-SEP-2000; 2000US-0229287.
01-SEP-2000; 2000US-0229343.
01-SEP-2000; 2000US-0229344.
01-SEP-2000; 2000US-0229345.
05-SEP-2000; 2000US-0229509.
05-SEP-2000; 2000US-0229513.
06-SEP-2000; 2000US-0230437.
06-SEP-2000; 2000US-0230438.
08-SEP-2000; 2000US-0231242.
08-SEP-2000; 2000US-0231243.
08-SEP-2000; 2000US-0231244.
08-SEP-2000; 2000US-0231413.
08-SEP-2000; 2000US-0231414.
08-SEP-2000; 2000US-0232080.
08-SEP-2000; 2000US-0232081.
12-SEP-2000; 2000US-0231968.
14-SEP-2000; 2000US-0232397.
14-SEP-2000; 2000US-0232398.
14-SEP-2000; 2000US-0232399.
14-SEP-2000; 2000US-0232400.
14-SEP-2000; 2000US-0232401.
14-SEP-2000; 2000US-0233063.
14-SEP-2000; 2000US-0233064.
14-SEP-2000; 2000US-0233065.
21-SEP-2000; 2000US-0234223.
21-SEP-2000; 2000US-0234274.
25-SEP-2000; 2000US-0234997.
25-SEP-2000; 2000US-0234998.
26-SEP-2000; 2000US-0235484.
27-SEP-2000; 2000US-0235834.
27-SEP-2000; 2000US-0235836.
29-SEP-2000; 2000US-0236327.
29-SEP-2000; 2000US-0236367.
29-SEP-2000; 2000US-0236368.
29-SEP-2000; 2000US-0236369.
29-SEP-2000; 2000US-0236370.
02-OCT-2000; 2000US-0236802.
02-OCT-2000; 2000US-0237037.
02-OCT-2000; 2000US-0237038.
02-OCT-2000; 2000US-0237039.
02-OCT-2000; 2000US-0237040.
13-OCT-2000; 2000US-0239935.
13-OCT-2000; 2000US-0239937.
20-OCT-2000; 2000US-0240960.
20-OCT-2000; 2000US-0241221.
20-OCT-2000; 2000US-0241785.
20-OCT-2000; 2000US-0241786.
20-OCT-2000; 2000US-0241787.
20-OCT-2000; 2000US-0241808.
20-OCT-2000; 2000US-0241809.
20-OCT-2000; 2000US-0241826.
01-NOV-2000; 2000US-0244617.
08-NOV-2000; 2000US-0246474.
08-NOV-2000; 2000US-0246475.
08-NOV-2000; 2000US-0246476.
08-NOV-2000; 2000US-0246477.
08-NOV-2000; 2000US-0246478.
08-NOV-2000; 2000US-0246523.
08-NOV-2000; 2000US-0246524.
08-NOV-2000; 2000US-0246525.
08-NOV-2000; 2000US-0246526.
08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246609.
08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249266.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX Disclosure; SEQ ID NO 34893; 307lpp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
XX Sequence 11363 BP; 3161 A; 2317 C; 2316 G; 3569 T; 0 other;

Query Match 3.6%; Score 34.2; DB 22; Length 11363;
Best Local Similarity 52.4%; Pred. No. 8.4;
Matches 75; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 720 AGTACTACTATTATCTTTTCAACGAGGCAACAAATCATACGACCAATGGAGAA 779

Db 9203 ATTGCTATTTTATTGTTATTTTAAATCTAAAGTTTATGGTTCCAGGGACUTTT 9262
QY 780 TTTGGGAGCTACATGGAAGGAAACCCCTAGAAATGGATCAGCAGGTGTGAAGGAGTACGAG 839
Db 9263 TATTGACCCCAACACACAGTATCCACAGGATATGGAGGCAAGGGACAAAGAGCATGAG 9322
QY 840 TTTTTCGGGGAAATATGCTGA 862
Db 9323 TCTTTTCCAGAAACTGCTGA 9345
RESULT 14
AAK80079
ID AAK80079 standard; DNA; 11366 BP.
XX
AC AAK80079;
XX
DT 07-NOV-2001 (first entry)
XX Human immune/hematopoietic antigen genomic sequence SEQ ID NO:34891.
XX Human; immune; haematopoietic; immune/hematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0203515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.

PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249219.
PR	17-NOV-2000;	2000US-0249224.
PR	17-NOV-2000;	2000US-0249235.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249300.
PR	17-NOV-2000;	2000US-0249309.
PR	01-DEC-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0251988.
PR	06-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254907.
PR	05-JAN-2001;	2001US-0256978.
XX	{HUMA-}	HUMAN GENOME SCI INC.

WPI: 2001-483426/52.

Disclosure; SEQ ID NO 34891; 3071pp + Sequence Listing; English, English

Sequence 11366 BP; 3164 A; 2317 C; 2316 G; 3569 T; 0 other;

Best Local Similarity 52.48; Pred. No. 8.4;

Matches	75;	Conservative	0;	Mismatches	68;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

720 AGTACTACTATTATCCTTATTCAACGAGGCAACAATCATACGGGACCAATGGAGGAA 779

9206 ATTGCTATTTTATTGTTATTTTAAACATGAATCTAAAGTTTATGTTCCAGGACTTTT 9265

780 TTTGGAGCTACATGGAAGGAACCTAGAAATGGATCAGGAGGTGAAGGAGTACGAG

9266 TATTGACCCACACACAGTATCCCAGGATATGGAGGCAAGGGGAACAAAGAGCATGAG 9325

840 TTTTTCGGGGAATATGGTGA 862

9326 TCTTTTCCAAGAAACTGGTGA 9348

RESULT 15

AAK80050	
ID	AAK80080 standard; DNA; 11366 BP.
XX	
AC	AAK80080;
XX	
DT	07-NOV-2001 (first entry)
XX	
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34892.
XX	
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW	cytostatic; gene therapy; vaccine; metastasis; ds.
XX	
OS	Homo sapiens.
PN	WO200157182-A2.
XX	
PD	09-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01354.
XX	
PF	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225266.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225268.
PR	14-AUG-2000; 2000US-0225270.
PR	14-AUG-2000; 2000US-0225447.
PR	14-AUG-2000; 2000US-0225757.
PR	14-AUG-2000; 2000US-0225758.
PR	14-AUG-2000; 2000US-0225759.
PR	18-AUG-2000; 2000US-0226279.
PR	22-AUG-2000; 2000US-0226661.
PR	22-AUG-2000; 2000US-0226668.
PR	22-AUG-2000; 2000US-0227182.
PR	23-AUG-2000; 2000US-0227009.
PR	30-AUG-2000; 2000US-0228924.
PR	01-SEP-2000; 2000US-0229287.
PR	01-SEP-2000; 2000US-0229343.
PR	01-SEP-2000; 2000US-0229344.
PR	01-SEP-2000; 2000US-0229345.
PR	05-SEP-2000; 2000US-0229509.
PR	05-SEP-2000; 2000US-0229513.
PR	06-SEP-2000; 2000US-0230437.
PR	06-SEP-2000; 2000US-0230438.
PR	08-SEP-2000; 2000US-0231242.
PR	08-SEP-2000; 2000US-0231243.
PR	08-SEP-2000; 2000US-0231244.
PR	08-SEP-2000; 2000US-0231413.
PR	08-SEP-2000; 2000US-0231414.
PR	08-SEP-2000; 2000US-0232080.
PR	08-SEP-2000; 2000US-0232081.
PR	12-SEP-2000; 2000US-0231968.
PR	14-SEP-2000; 2000US-0232398.
PR	14-SEP-2000; 2000US-0232399.
PR	14-SEP-2000; 2000US-0232400.
PR	14-SEP-2000; 2000US-0232401.
PR	14-SEP-2000; 2000US-0233063.
PR	14-SEP-2000; 2000US-0233064.
PR	14-SEP-2000; 2000US-0233065.
PR	21-SEP-2000; 2000US-0234223.
PR	21-SEP-2000; 2000US-0234274.
PR	25-SEP-2000; 2000US-0234997.
PR	25-SEP-2000; 2000US-0234998.
PR	26-SEP-2000; 2000US-0235484.
PR	27-SEP-2000; 2000US-0235834.
PR	27-SEP-2000; 2000US-0235836.
PR	29-SEP-2000; 2000US-0236327.
PR	29-SEP-2000; 2000US-0236367.
PR	29-SEP-2000; 2000US-0236368.
PR	29-SEP-2000; 2000US-0236369.
PR	29-SEP-2000; 2000US-0236370.
PR	02-OCT-2000; 2000US-0236802.
PR	02-OCT-2000; 2000US-0237037.
PR	02-OCT-2000; 2000US-0237038.
PR	02-OCT-2000; 2000US-0237039.
PR	02-OCT-2000; 2000US-0237040.
PR	13-OCT-2000; 2000US-0239937.
PR	13-OCT-2000; 2000US-0239938.
PR	20-OCT-2000; 2000US-0240960.
PR	20-OCT-2000; 2000US-0241221.
PR	20-OCT-2000; 2000US-0241785.
PR	20-OCT-2000; 2000US-0241786.
PR	20-OCT-2000; 2000US-0241808.
PR	20-OCT-2000; 2000US-0241809.
PR	20-OCT-2000; 2000US-0241826.
PR	01-NOV-2000; 2000US-0244617.
PR	08-NOV-2000; 2000US-0246474.
PR	08-NOV-2000; 2000US-0246475.
PR	08-NOV-2000; 2000US-0246476.
PR	08-NOV-2000; 2000US-0246477.
PR	08-NOV-2000; 2000US-0246478.
PR	08-NOV-2000; 2000US-0246523.
PR	08-NOV-2000; 2000US-0246524.
PR	08-NOV-2000; 2000US-0246525.
PR	08-NOV-2000; 2000US-0246526.
PR	08-NOV-2000; 2000US-0246527.
PR	08-NOV-2000; 2000US-0246528.
PR	08-NOV-2000; 2000US-0246532.
PR	08-NOV-2000; 2000US-0246609.
PR	08-NOV-2000; 2000US-0246610.
PR	08-NOV-2000; 2000US-0246611.
PR	08-NOV-2000; 2000US-0246613.
PR	17-NOV-2000; 2000US-0249207.
PR	17-NOV-2000; 2000US-0249208.
PR	17-NOV-2000; 2000US-0249209.
PR	17-NOV-2000; 2000US-0249210.
PR	17-NOV-2000; 2000US-0249211.
PR	17-NOV-2000; 2000US-0249212.
PR	17-NOV-2000; 2000US-0249213.
PR	17-NOV-2000; 2000US-0249214.
PR	17-NOV-2000; 2000US-0249215.
PR	17-NOV-2000; 2000US-0249216.
PR	17-NOV-2000; 2000US-0249217.
PR	17-NOV-2000; 2000US-0249218.
PR	17-NOV-2000; 2000US-0249244.
PR	17-NOV-2000; 2000US-0249245.
PR	17-NOV-2000; 2000US-0249264.
PR	17-NOV-2000; 2000US-0249265.
PR	17-NOV-2000; 2000US-0249297.
PR	17-NOV-2000; 2000US-0249299.
PR	17-NOV-2000; 2000US-0249300.
PR	01-DEC-2000; 2000US-0250160.
PR	01-DEC-2000; 2000US-0250191.

Imis Page Blank (uspto)

M nucleic - nucleic search, using sw model

run on: January 25, 2003, 21:06:06 ; Search time 58 Seconds
(without alignments)
4991.429 Million cell updates/sec

title: US-09-701-023-1_COPY_81_1024

effect score: 944
sequence: 1 tggcgactctctctctc.....attgtcttggaagctttaa 944

scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0.

sarched: 441362 seqs, 153338381 residues

total number of hits satisfying chosen parameters: 882724

imum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

itabase : Issued Patents.NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result	No.	Score	Query Match	Length	DB ID	Description
1	45.8	4.9	7218	1	US-08-232-463-14	Sequence 14, Appl
2	36.8	3.9	7218	1	US-08-232-463-14	Sequence 14, Appl
3	34.2	3.6	494	2	US-08-332-766A-22	Sequence 22, Appl
4	32.8	3.5	1611	1	US-08-485-569-1	Sequence 1, Appl
5	32.8	3.5	1611	1	US-08-480-993-1	Sequence 1, Appl
6	32.8	3.5	1611	2	US-07-903-079B-1	Sequence 1, Appl
7	32.8	3.5	2894	1	US-08-278-091-1	Sequence 1, Appl
8	32.8	3.5	2894	1	US-08-483-859-1	Sequence 1, Appl
9	32.8	3.5	2894	1	US-08-472-173-1	Sequence 1, Appl
10	32.8	3.5	2894	2	US-08-487-167-1	Sequence 1, Appl
11	32.8	3.5	2894	2	US-08-482-816-1	Sequence 1, Appl
12	32.8	3.5	2894	2	US-08-296-149-1	Sequence 1, Appl
13	32.8	3.5	2894	2	US-08-801-439-1	Sequence 1, Appl
14	32.8	3.5	2894	2	US-08-615-271-1	Sequence 1, Appl
15	32.8	3.5	2894	3	US-09-074-660-1	Sequence 1, Appl
16	32.8	3.5	2894	3	US-09-074-659-1	Sequence 1, Appl
17	32.8	3.5	2894	3	US-09-106-468-1	Sequence 1, Appl
18	32.8	3.5	2894	3	US-09-106-466A-1	Sequence 1, Appl
19	32.8	3.5	2894	3	US-09-106-467-1	Sequence 1, Appl
20	32.6	3.5	1631	3	US-09-118-319-1	Sequence 1, Appl
21	32.2	3.4	1925	2	US-08-553-436A-1	Sequence 1, Appl
22	32	3.4	243	1	US-07-922-723A-9	Sequence 9, Appl
23	32	3.4	243	1	US-07-799-828C-9	Sequence 9, Appl
24	32	3.4	243	1	US-08-074-275-9	Sequence 9, Appl
25	32	3.4	243	1	US-08-480-366-9	Sequence 9, Appl
26	32	3.4	243	2	US-07-952-277A-9	Sequence 9, Appl
27	31.6	3.3	1798	4	US-08-687-691B-1	Sequence 1, Appl

Sequence 16, Appl
Sequence 1, Appl
Sequence 9, Appl
Sequence 35, Appl
Sequence 22, Appl
Sequence 22, Appl
Sequence 25, Appl
Sequence 25, Appl
Sequence 4, Appl
Sequence 9, Appl
Sequence 11, Appl
Sequence 12, Appl
Sequence 12, Appl
Sequence 7, Appl
Sequence 7, Appl
Sequence 70, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9pt-F15
US-08-232-463-14

Query Match 4.9%; Score 45.8; DB 1; Length 7218;

Best Local Similarity 8.5%; Pred. No. 9.3e-05;
Matches 17; Conservative 116; Mismatches 68; Indels 0; Gaps 0;

QY 96 GCGGAGATCGGACAGACACTGAAGAGTCGTGGTGGGAAACCCAGCATCGAAGACAGGT 155
Db 1236 RRR 1177
QY 156 CAGCAAAACAGAGAAACCAAGCTTGAGAGCAATGGGTCTAGCAAAAGCTCGAGCGTCAG 215
Db 1176 RRR 1117
QY 216 AGAATCGAAGAGAAAGCAACTCGCGCGGCCACAGTCGGAGACAGTCATCAGTA 275
Db 1116 RRR 1057
QY 276 GCATCGATCTTAACAACGCT 296
Db 1056 CCTCGACCTGCAGCAAGCT 1036

RESULT 2
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 28-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9pt-Fls
US-08-232-463-14

Query Match 3.9%; Score 36.8; DB 1; Length 7218;
Best Local Similarity 4.7%; Pred No 0.1

QY 380 TGGAGTCGGTCGGTCAGGTTATGATCCACCCGGTTATTTCTCCATGGGTTTGTTCGA 439
Db 1040 TGGCTGCAGTCGAGGAGCTTGGCATTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1099
QY 440 GACCTCCCTCCACTACTCATGAGCTCTCTTCAATCTCAATCTCAATCTCAATCTCA 499
Db 1100 YY 1159
QY 500 TTCCAATATCTGCTGACACTTGTCTCAAGAAGAAAGCTTTTGGATGGTGATCAGAATAA 559
Db 1160 YY 1219
QY 560 TGTAGTTCGATCCACGGTGGTGGATTTTGAATACACATGATCTCTCTCCGATGAA 619
Db 1220 YY 1279
QY 620 CGGTACGATCAGTATCTCTTCAATCAGATCATCATCAGAGGAGCCCAAGGTTTCTT 679
Db 1280 YY 1339
QY 680 TGTATCAGATCGCTAGAGCAGCTTCACTTTCTGCTTCTAGTACTACTTATTAATCCTTA 739
Db 1340 YY 1399
QY 740 TTTC 743
Db 1400 YYY 1403

RESULT 3
US-08-332-766A-22
; Sequence 22, Application US/08332766A
; Patent No. 5843647
; GENERAL INFORMATION:
; APPLICANT: JEFFREYS, Alec J.
; APPLICANT: ARMOUR, John
; TITLE OF INVENTION: SIMPLE TANDEM REPEATS
; NUMBER OF SEQUENCES: 125
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,766A
; FILING DATE: 01-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9326052.9
; FILING DATE: 21-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BIRD, Donald J.
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 217211/M94/0434/GB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 494 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

Db 968 TCAATGCTGATTTAGCCAAAGCCCTTTTAATGTAAGCGCGCAACAAGGTGCTATTTGTAAGTG 1027
Qy 790 ACATGGAAGGAACCCAGAAATGGATCAGGAGGTGTGAAGGAGTACGAGTTTTCGCGG 849
Db 1028 AAGTTTACCGAATCTGCTGCTGAAAGAGCAGGACTTAAAGCGGCGATATATATCAGG 1087
Qy 850 GGAATATGTTGAAGAGTTTCACTGCTGGTGCCTACAA 885
Db 1088 CGATGAACGGTCAAAAAATCTCAAGTTTCGCTGAAA 1123

RESULT 6
US-07-903-079B-1
; Sequence 1, Application US/07903079B
; Patent No. 5843463
; GENERAL INFORMATION:
; APPLICANT: Krivan, Howard C.
; APPLICANT: Samuel, James E.
; APPLICANT: No. 5843463berg, Nils T.
; TITLE OF INVENTION: ADHESIN-OLIGOSACCHARIDE CONJUGATE
; TITLE OF INVENTION: VACCINE FOR HAEMOPHILUS INFLUENZAE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/903,079B
FILING DATE: 22-JUN-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,966
FILING DATE: 20-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/631,698
FILING DATE: 21-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7969-004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1611 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-07-903-079B-1

Query Match 3.5%; Score 32.8; DB 2; Length 1611;
Best Local Similarity 50.6%; Pred. No. 0.89;
Matches 79; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
Qy 730 TTAATCCTTATTTCAAGCAGGCAACAATCATACGGGACCAATGGAGCAATTTGGAGCT 789
Db 968 TCAATGCTGATTTAGCCAAAGCCCTTTTAATGTAAGCGCGCAACAAGGTGCTAAGTG 1027
Qy 790 ACATGGAAGGAACCCCTAGAAATGGATCAGGAGGTGTGAAGGAGTACGAGTTTTCGCG 849

Qy 850 GGAATATGTTGAAGAGTTTCACTGCTGGTGCCTACAA 885
Db 1088 CGATGAACGGTCAAAAAATCTCAAGTTTCGCTGAAA 1123

RESULT 7
US-08-278-091-1
; Sequence 1, Application US/08278091
; Patent No. 5506139
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/278,091
FILING DATE: 21-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-371
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-278-091-1

Query Match 3.5%; Score 32.8; DB 1; Length 2894;
Best Local Similarity 50.6%; Pred. No. 1.3;
Matches 79; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
Qy 730 TTAATCCTTATTTCAAGCAGGCAACAATCATACGGGACCAATGGAGCAATTTGGAGCT 789
Db 1655 TCAATGCTGATTTAGCCAAAGCCCTTTTAATGTAAGCGCGCAACAAGGCGCATTTGTAAGTG 1714
Qy 790 ACATGGAAGGAACCCCTAGAAATGGATCAGGAGGTGTGAAGGAGTACGAGTTTTCGCG 849
Db 1715 AAGTTTACCGAATCTGCTGCTGAAAGAGCAGGACTTAAAGCGCGGATATATATCAGG 1774

Qy 850 GGAATATGTTGAAGAGTTTCACTGCTGGTGCCTACAA 885
Db 1775 CGATGAACGGTCAAAAAATCTCAAGTTTCGCTGAAA 1810

RESULT 8
US-08-483-859-1
; Sequence 1, Application US/08483859
; Patent No. 5656436
; GENERAL INFORMATION:

APPLICANT: CHONG, Pele
APPLICANT: OOMEN, Raymond P.
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
TITLE OF INVENTION: Reduced Protease Activity
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/483,859
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-495 MIS:vg
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
S-08-483-859-1
Query Match 3.5%; Score 32.8; DB 1; Length 2894;
Best Local Similarity 50.6%; Pred. No. 1.3;
Matches 79; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

y 730 TTAATCCTTATTTCACAGGCGCAACAAATCATACGGGACCAATGGAGGAAATTTGGGAGCT 789
|||||
b 1655 TCAATGCTGATTTAGCCAAAGCCTTTAATGTAAGCGCGCAACAAGCGCATTTGTAAGTG 1714
y 790 ACATGGAAGAAACCCCTAGAAATGGATCAGGAGGTGTGAAGGAGTACGAGTATTTTCCGG 849
|||||
b 1715 AAGTTTACCGAAATCTGCTGCTGAAAAAGCAGGACTTAAAGCGGCGATATTATCACGG 1774
y 850 GGAATATGCTGAAGAGTTTCAGTGTGGCTACAA 885
|||||
b 1775 CGATGACGGTCAAAAATCTCAAGTTTCGCTGAAA 1810

RESULT 9
S-08-472-173-1
Sequence 1, Application US/08472173
Patent No. 5665353
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M
APPLICANT: YANG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: OOMEN, Raymond P.
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
TITLE OF INVENTION: Reduced Protease Activity
NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/472,173
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-493 MIS:vg
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-472-173-1
Query Match 3.5%; Score 32.8; DB 1; Length 2894;
Best Local Similarity 50.6%; Pred. No. 1.3;
Matches 79; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 730 TTAATCCTTATTTCACAGGCGCAACAAATCATACGGGACCAATGGAGGAAATTTGGGAGCT 789
|||||
Db 1655 TCAATGCTGATTTAGCCAAAGCCTTTAATGTAAGCGCGCAACAAGCGCATTTGTAAGTG 1714
QY 790 ACATGGAAGAAACCCCTAGAAATGGATCAGGAGGTGTGAAGGAGTACGAGTATTTTCCGG 849
|||||
Db 1715 AAGTTTACCGAAATCTGCTGCTGAAAAAGCAGGACTTAAAGCGGCGATATTATCACGG 1774
QY 850 GGAATATGCTGAAGAGTTTCAGTGTGGCTACAA 885
|||||
Db 1775 CGATGACGGTCAAAAATCTCAAGTTTCGCTGAAA 1810

RESULT 10
US-08-487-167-1
Sequence 1, Application US/08487167
Patent No. 5669302
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: YANG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: OOMEN, Raymond P.
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
TITLE OF INVENTION: Reduced Protease Activity
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada

```

; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,167
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-508 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-487-167-1

```

```

Query Match 3.5%; Score 32.8; DB 2; Length 2894;
Best Local Similarity 50.6%; Pred. No. 1.3;
Matches 79; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 730 TTAATCCTTATTTCACGAGGCAACAAATCATACGGGACCAATGGAGGAATTTGGGAGCT 789
D 1655 TCAATGCTGATTAGCCAAAGCCTTTAATGTAAGCGCGCAACAAAGCGCATTTGTAAGTG 1714

QY 790 ACATGGAAGAAACCCCTAGAAATGGATCAGGAGGTGTGAAGGAGTACGAGTTCCTCCGG 849
D 1715 AAGTTTACCGAATCTGCTGCTGAAAGGACGAGCTTAAAGCGGCGATATTATCACGG 1774

QY 850 GGAATATGCTGAAAGAGTTTCAGTGGTGCTACAA 885
D 1775 CGATGAACGCTCAAAAATCTCAAGTTTCGCTGAAA 1810

```

```

RESULT 11
US-08-482-816-1
; Sequence 1, Application US/08482816
; Patent No. 5935573
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease Act
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; APPLICATION NUMBER: US/08/482,816
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-494 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-482-816-1

```

```

Query Match 3.5%; Score 32.8; DB 2; Length 2894;
Best Local Similarity 50.6%; Pred. No. 1.3;
Matches 79; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 730 TTAATCCTTATTTCACGAGGCAACAAATCATACGGGACCAATGGAGGAATTTGGGAGCT 789
D 1655 TCAATGCTGATTAGCCAAAGCCTTTAATGTAAGCGCGCAACAAAGCGCATTTGTAAGTG 1714

QY 790 ACATGGAAGAAACCCCTAGAAATGGATCAGGAGGTGTGAAGGAGTACGAGTTCCTCCGG 849
D 1715 AAGTTTACCGAATCTGCTGCTGAAAGGACGAGCTTAAAGCGGCGATATTATCACGG 1774

QY 850 GGAATATGCTGAAAGAGTTTCAGTGGTGCTACAA 885
D 1775 CGATGAACGCTCAAAAATCTCAAGTTTCGCTGAAA 1810

```

```

RESULT 12
US-08-296-149-1
; Sequence 1, Application US/08296149
; Patent No. 5939297
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,149
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:

```

REFERENCE/DOCKET NUMBER: 1038-390

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2894 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IS-08-296-149-1

Query Match 3.5%; Score 32.8; DB 2; Length 2894;

Best Local Similarity 50.6%; Pred. No. 1.3;

Matches 79; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

y 730 TTAATCCTTATTTCAAGGAGGCAACAAATCATACGGACCAATGGAGGAGT 789

b 1655 TCAATGCTGATTTAGCCAAAGCCTTTAATGTAAGCGCGCAACAGCGCATTTGTAAGTG 1714

y 790 ACATGAAGGAACCCCTAGAAATGGATCAGAGGTGTGAAGGAGTACGAGT 849

b 1715 AAGTTTACCGAAATCTGCTGCTGTAAGAGCAGGACTTAAGCGGGATATTATCACGG 1774

y 850 GGAATATGTTGAAGAGTTTCAGTGGTGCTACAA 885

b 1775 CGATGACGTCATAAAATCTCAAGTTTCGCTGAAA 1810

RESULT 13

S-08-801-499-1

Sequence 1, Application US/08801499

Patent No. 5962430

GENERAL INFORMATION:

APPLICANT: LOOSMORE, Sheena M

APPLICANT: YANG, Yan-Ping

APPLICANT: CHONG, Pele

APPLICANT: OOMEN, Raymond P.

APPLICANT: KLEIN, Michel H.

TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease Act

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney

STREET: 6th Floor, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/801,499

FILING DATE:

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/482,816

FILING DATE: 07-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/296,149

FILING DATE: 26-AUG-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/278,091

FILING DATE: 21-JUL-1994

ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-671 MIS:jb

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2894 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-801-499-1

Query Match 3.5%; Score 32.8; DB 2; Length 2894;

Best Local Similarity 50.6%; Pred. No. 1.3;

Matches 79; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 730 TTAATCCTTATTTCAACGAGGCAACAAATCATACGGACCAATGGAGGAGCT 789

Db 1655 TCAATGCTGATTTAGCCAAAGCCTTTAATGTAAGCGCGCAACAGCGCATTTGTAAGTG 1714

QY 790 ACATGAAGGAACCCCTAGAAATGGATCAGGAGGTGTGAAGGAGTACGAGT 849

Db 1715 AAGTTTACCGAAATCTGCTGCTGTAAGAGCAGGACTTAAGCGGGATATTATCACGG 1774

QY 850 GGAATATGTTGAAGAGTTTCAGTGGTGCTACAA 885

Db 1775 CGATGACGTCATAAAATCTCAAGTTTCGCTGAAA 1810

RESULT 14

US-08-615-271-1

Sequence 1, Application US/08615271

Patent No. 5981503

GENERAL INFORMATION:

APPLICANT: LOOSMORE, Sheena M.

APPLICANT: YANG, Yan-Ping

APPLICANT: CHONG, Pele

APPLICANT: OOMEN, Raymond P.

APPLICANT: KLEIN, Michel H.

TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney

STREET: 6th Floor, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/615,271

FILING DATE: 20-JUN-1996

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-580

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2894 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-615-271-1

Query Match 3.5%; Score 32.8; DB 2; Length 2894;

Best Local Similarity 50.6%; Pred. No. 1.3;

Matches 79; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

M nucleic - nucleic search, using sw model
an on: January 25, 2003, 22:30:57 ; Search time 63 Seconds
(without alignments)
6731.841 Million cell updates/sec

itle: US-09-701-023-1_COPY_81_1024
arfect score: 944
sequence: 1 tggcgactctctctcttc.....atttgtcttgaagcttta 944

oring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

arched: 396772 seqs, 224632407 residues
otal number of hits satisfying chosen parameters: 793544

inimum DB seq length: 0
aximum DB seq length: 2000000000
st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : Published Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %	Score	Match	Length	DB ID	Description
1	35.8	3.8	465237	10	US-09-933-267A-1	Sequence 1, Appli
2	34.6	3.7	2269	9	US-09-989-920-113	Sequence 113, App
3	34.2	3.6	389	9	US-09-796-692-2664	Sequence 2664, Ap
4	34	3.6	470	10	US-09-864-761-701	Sequence 701, App
5	33.6	3.6	307	10	US-09-867-701-7516	Sequence 7516, Ap
6	33	3.5	265	10	US-09-764-864-753	Sequence 753, App
7	33	3.5	331	10	US-09-864-761-29232	Sequence 29232, A
8	33	3.5	365	10	US-09-764-864-338	Sequence 338, App
9	33	3.5	511	10	US-09-864-761-12665	Sequence 12665, A
10	32.8	3.5	1381	9	US-09-938-842A-4232	Sequence 4232, Ap
11	32.4	3.4	286	10	US-09-867-701-8693	Sequence 8693, Ap
12	32.4	3.4	387	10	US-09-864-761-16334	Sequence 16334, A
13	32.4	3.4	470	10	US-09-864-761-1484	Sequence 1484, Ap
14	32.4	3.4	470	10	US-09-864-761-15622	Sequence 15622, A
15	32.2	3.4	1038	9	US-09-864-921-11	Sequence 11, Appli
16	32.2	3.4	1101	10	US-09-815-242-4645	Sequence 4645, Ap
17	32.2	3.4	1107	9	US-09-864-921-84	Sequence 84, Appli
18	32.2	3.4	1155	10	US-09-815-242-8511	Sequence 8511, Ap
19	32.2	3.4	3111	10	US-09-728-721-56	Sequence 56, Appli

c 20 32.2 3.4 4244 10 US-09-728-721-54 Sequence 54, Appli
c 21 32 3.4 2000 9 US-09-938-842A-2854 Sequence 2854, Ap
c 22 32 3.4 32768 10 US-09-070-927A-100 Sequence 100, App
c 23 31.6 3.3 1798 8 US-08-687-691B-1 Sequence 1, Appli
c 24 31.6 3.3 148567 10 US-09-801-876B-3 Sequence 3, Appli
c 25 31.4 3.3 270 10 US-09-294-093B-4115 Sequence 4115, Ap
c 26 31.4 3.3 300 10 US-09-294-093B-5388 Sequence 5388, Ap
c 27 31.4 3.3 2000 9 US-09-938-842A-4161 Sequence 4161, Ap
c 28 31.4 3.3 13605 10 US-09-764-877-3446 Sequence 3446, Ap
c 29 31.4 3.3 32768 10 US-09-070-927A-17 Sequence 17, Appli
c 30 31.2 3.3 356 9 US-09-796-692-8787 Sequence 8787, Ap
c 31 31.2 3.3 368004 10 US-09-949-654-3 Sequence 3, Appli
c 32 31 3.3 2968 10 US-09-822-849A-301 Sequence 301, App
c 33 31 3.3 7805 12 US-10-044-090-370 Sequence 370, App
c 34 31 3.3 11872 10 US-09-764-847-1390 Sequence 1390, Ap
c 35 31 3.3 152331 9 US-10-095-407-16 Sequence 16, Appli
c 36 30.8 3.3 2815 10 US-09-764-855-301 Sequence 301, App
c 37 30.8 3.3 2851 10 US-09-764-855-300 Sequence 300, App
c 38 30.8 3.3 2851 10 US-09-764-855-302 Sequence 302, App
c 39 30.8 3.3 174424 10 US-09-967-768A-314 Sequence 314, App
c 40 30.6 3.2 174493 10 US-09-804-471A-3 Sequence 3, Appli
c 41 30.4 3.2 621 10 US-09-350-756-11 Sequence 11, Appli
c 42 30.4 3.2 684 10 US-09-867-701-5214 Sequence 5214, Ap
c 43 30.4 3.2 1323 9 US-09-910-186A-3 Sequence 3, Appli
c 44 30.4 3.2 1326 9 US-09-910-186A-5 Sequence 5, Appli
c 45 30.4 3.2 1332 9 US-09-910-186A-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-933-267A-1/c
; Sequence 1, Application US/099333267A
; Patent No. US20020123095A1
; GENERAL INFORMATION:
; APPLICANT: Kalush, Francis et al.
; TITLE OF INVENTION: Estrogen receptor alpha variants and
; FILE REFERENCE: methods of detection thereof
; FILE REFERENCE: C1000258C14
; CURRENT APPLICATION NUMBER: US/09/933,267A
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/160626
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 60/183756
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: 09/692414
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/768184
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 09/804076
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 09/826314
; PRIOR FILING DATE: 2001-04-05
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 465237
; TYPE: DNA
; ORGANISM: human
US-09-933-267A-1

Query Match 3.8%; Score 35.8; DB 10; Length 465237;
Best Local Similarity 52.9%; Pred. No. 7;
Matches 99; Conservative 0; Mismatches 87; Indels 1; Gaps 1;
QY 97 CGGAGATCGGCACAGACACTGACAGATCGTGGTCGGAACCCAGGATCGAGAGAGGTC 156
Db 2868 GAGACACAG 2810
QY 157 AGCAAAACACAG 216
Db 2809 AGCAAG 2750

QY	217	GAATCGAAGAAGAAAAGCAACTCGCGCCGACACAGTCGGAGACACGCATCAGTAG	276
Db	2749	GAAAGAAAGAAAAGAAAAGGAAGGAAAAAGAAAAGAACATAAATCTTCTCATGAC	2690
QY	277	CATCGAT	283
Db	2689	ATTGGAT	2683

```

RESULT 2
US-09-989-920-113
; Sequence 113, Application US/099899320
; Patent No. US20020172957A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Protein
; FILE REFERENCE: DEX-0291
; CURRENT APPLICATION NUMBER: US/09/989,920
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/252,500
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 113
; LENGTH: 2269
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2266)..(2266)
; OTHER INFORMATION: a, c, g or t
US-09-989-920-113

```

[illegible]

[illegible]

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 12665
LENGTH: 511
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC015897.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.94
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.82
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.94
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.84
US-09-854-761-12665

Query Match 3.5%; Score 33; DB 10; Length 511;
Best Local Similarity 53.5%; Pred. No. 1;
Matches 69; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

y 106 GCACAGACACTGCTGAAGAGTGGTGTCGCGAAACCAGGATCGAAGACAGGTCCAGCAAAAAC 165
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
yb 79 GGGCCTTGACAAAAAACCGTCTTGC CGAATCTAGGAAGAAGAAGACGACCTGAGNAAGG 138
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
y 166 AGAAGAAACCACGTTGTGAGAGGAATGGGTGTAGCAAAGCTCGAGCGTCAGAGTAATCGAAG 225
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
yb 139 AGAGGAACCTGAAGAAGGCTGAGGGGAAAGTAGGAAAGAGAGAGAGAGTATGTGATACCAAA 198
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
y 226 AAGAAAGA 234
||||| ||
yb 199 AAGAAATGA 207
||||| ||

RESULT 10
US-09-938-842A-4232/c
Sequence 4232, Application US/09938842A
Patent NO. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SCRP1300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 4232
LENGTH: 1381
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-4232

Query Match 3.5%; Score 32.8; DB 9; Length 1381;
Best Local Similarity 47.2%; Pred No. 2.2;
Matches 100; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

y 397 AGGTATGATGACCCGGTTATTTCCTCCATGGGTTTTGTTGAGAGCCTCCTCCACTACTC 456
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

M protein - protein search, using sw model

Run on: January 25, 2003, 22:37:22 ; Search time 70 Seconds
(without alignments)
597.724 Million cell updates/sec

Title: US-09-701-023-4

Perfect score: 1634

Sequence: 1 MATSLFFSTQNSVGNPNPND.....TTSSLVGCSPNTLDLSKL 314

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1628	99.6	314	21	Arabidopsis thalia
2	104	6.4	575	22	Drosophila melanog
3	103	6.3	1693	22	Drosophila melanog
4	102	6.2	430	22	Novel human diagno
5	102	6.2	430	22	Novel human diagno
6	102	6.2	631	22	Drosophila melanog
7	102	6.2	1406	22	Drosophila melanog
8	101.5	6.2	729	22	Drosophila melanog
9	100	6.1	1059	22	Drosophila melanog
10	98.5	6.0	607	22	Drosophila melanog

11	98	6.0	603	22	ABB68633	Drosophila melanog
12	96.5	5.9	669	22	AAE01962	Arabidopsis thalia
13	96.5	5.9	704	22	ABB61723	Drosophila melanog
14	96.5	5.9	704	22	ABB61756	Drosophila melanog
15	96.5	5.9	1098	22	ABG13249	Novel human diagno
16	96	5.9	751	22	ABB58312	Drosophila melanog
17	95.5	5.8	697	22	ABB64801	Drosophila melanog
18	95.5	5.8	1165	22	AAE07091	S cerevisiae apopt
19	95	5.8	157	22	ABG08424	Novel human diagno
20	95	5.8	625	20	AAV08487	S. cerevisiae PacC
21	94.5	5.8	337	21	AAG12715	Arabidopsis thalia
22	94.5	5.8	349	21	AAG12714	Arabidopsis thalia
23	94.5	5.8	417	21	AAG12713	Arabidopsis thalia
24	94	5.8	461	21	AAE58247	Arabidopsis thalia
25	94	5.8	1412	22	AAE59278	Human SRC-3 protei
26	93	5.7	359	20	AAV05667	Maize cinnamyl alc
27	93	5.7	655	23	ABB57126	Mouse ischaemic co
28	93	5.7	656	14	AAE44555	Human EWS protein
29	93	5.7	656	19	AAE38133	Human EWS protein
30	93	5.7	665	22	ABG06460	Novel human diagno
31	92.5	5.7	506	22	ABB63474	Drosophila melanog
32	92.5	5.7	1455	22	ABB63040	Drosophila melanog
33	92.5	5.7	3080	22	ABB64877	Drosophila melanog
34	92	5.6	922	21	AAE36515	Candida albicans C
35	92	5.6	1116	22	ABG20642	Novel human diagno
36	92	5.6	1464	19	AAE42632	Human transcriptio
37	92	5.6	2444	22	ABB62786	Drosophila melanog
38	91.5	5.6	712	23	ABB97934	Human protein sequ
39	91.5	5.6	837	22	ABB58272	Drosophila melanog
40	91.5	5.6	996	22	ABB65591	Drosophila melanog
41	91.5	5.6	1289	22	ABB70840	Drosophila melanog
42	91	5.6	182	23	ABP33230	Human ORF2203 prot
43	91	5.6	734	20	AAV14204	Soybean PvAlf tran
44	91	5.6	951	20	AAE81028	ER interacting dom
45	91	5.6	1420	20	AAE61025	AB1 (Amplified in

ALIGNMENTS

RESULT 1
AAB26111 standard; Protein; 314 AA.
ID AAB26111 standard; Protein; 314 AA.
XX AAB26111;
XX 30-JAN-2001 (first entry)
DT Arabidopsis thaliana sporocyteless spl protein.
DE Arabidopsis thaliana sporocyteless spl protein.
XX Sporocyteless; spl; meiocyte formation; plant sterility; seedless fruit;
KW pollenless flower; Ds element.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
FH Key Location/Qualifiers
FT Misc-difference 23 /label= unknown
FT FT /note= "shown as Lrg, encoded by AGA"
XX W0200055907-A1.
XX 28-SEP-2000.
XX 22-MAR-1999; 99WO-SG00023.
XX 22-MAR-1999; 99WO-SG00023.
XX (MOLE-) INST MOLECULAR AGROBIOLOGY.
XX Ye D, Yang W, Sundaresan V, Xu J;
XX WPI: 2000-594578/56.
DR

DR	N-PSDB; AAA94065.
XX	
PT	Sporocyteless nucleic acids and polypeptides, useful for controlling sporocyte or melocyte formation in plants for producing transgenics that bear seedless fruits and/or pollenless flowers -
XX	
PS	Claim 10; Fig 3; 65pp; English.
XX	
CC	The present sequence is the Arabidopsis thaliana sporocyteless (SPL) protein. This protein is involved in the formation of microsporocytes in male plants and megasporocytes in female plants. The coding sequence for the mutant gene was isolated from a collection of transposants due to its male and female sterile phenotype. The mutation is caused by the insertion of the Ds element between bases 411 and 412 of the gene, causing a 4 base pair duplication. The identification of the sterile phenotype in plants caused by the insertion of this element means that it can be used to produce transgenic plants which contain the Ds element or antisense sequences to the gene, enabling the production of seedless fruits, pollenless flowers and plants with a larger biomass.
XX	
QQ	Sequence 314 AA;
Query Match	99.6%; Score 1628; DB 21; Length 314;
Best Local Similarity	99.7%; Pred. No. 5.6e-154;
Matches 31;	Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy	1 MATSLFFMSTDONSVGNPNDLRLNTRLVVNSSGEIETETLKSRRKPGSKTGQOKKPT 60
Dd	1 MATSLFFMSTDONSVGNPNDLXNTRLVVNSSGEIETETLKSRRKPGSKTGQOKKPT 60
Qy	61 LRGMGVAKLERQRIEEKKQLAAATVGTDSVASISNNATRLPVPDPGVVLQGFPSILG 120
Dd	61 LRGMGVAKLERQRIEEKKQLAAATVGTDSVASISNNATRLPVPDPGVVLQGFPSILG 120
Qy	121 SNRIYCGVGSGQWIMDVISFPGVFETSSTHELSSISNPOMFNASSNRCDTCFKKR 180
Dd	121 SNRIYCGVGSGQWIMDVISFPGVFETSSTHELSSISNPOMFNASSNRCDTCFKKR 180
Qy	181 LDGDQNVVRNSGGGFSKYTWIPPMNGDYQYLQSDHQHSQGFLYDHRTARAASVSAS 240
Dd	181 LDGDQNVVRNSGGGFSKYTWIPPMNGDYQYLQSDHQHSQGFLYDHRTARAASVSAS 240
Qy	241 STTINPYFNEATNHITGPMEEFGSYMEGNPRNGSGGVKEYEFPFGYGERVSVATTSSLV 300
Dd	241 STTINPYFNEATNHITGPMEEFGSYMEGNPRNGSGGVKEYEFPFGYGERVSVATTSSLV 300
Qy	301 GDGSPNTIDLSKL 314
Dd	301 GDGSPNTIDLSKL 314

RESULT 2
ABB70882
ID ABB70882 standard; Protein; 575 AA.

```

PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX N-PSDB; ABL14985.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 39438; 2lpp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL3051), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 575 AA;
SQ
Query Match          6.4%; Score 104; DB 22; Length 575;
Best Local Similarity 23.5%; Pred. No. 0.24;
Matches 55; Conservative 31; Mismatches 86; Indels 62; Gaps 9;
Qy 43 RRRPGSKTGOKKPTLRGMGVAKLERQRIEEKKOLAAATVGDTSVASISNNATRL 102
   ||| : ||| : :||| : :||| : :||| :
Db 27 RGRRCSDVQQOQAATCHRRW-----VKRNRIQD-----ASLGSSDDDEFLLG---- 69
Qy 103 PVPVDPVVLQGFSSSLGSNRIYCGVGCSGOVMIDPVISPNGFVETSTTHLESSLNPQ 162
   ||| ||| : :||| : :||| : :||| : :||| :
Db 70 -----VLRG-PSTFANAFLY---VGLGTVALGLVIAFVTGTEKGPKFTTELRLIGPSL 117
Qy 163 MFNASNNRCDTCFKKKR-----LDGQNNVVRNSGGCFSKYTMIPPMNGDYQLIQS 216
   : ||| : ||| : :||| : :||| : :||| :
Db 118 IGNKIANGCKQLNIKKTRKKNKIDADHTTLLRNESKRKPYPYA----- 163
Qy 217 DHQRSGFGLYDHRIARAASVSASS-----TTIN---PYFNETHNHGPME 259
   ||| : ||| : :||| : :||| : :||| : :||| :
Db 164 --HRRKSQMLNEGMEALROIATTSLFMONEOKTAIRNWVPLINEPEGSDAPLE 215

```

RESULT 3	
ABB64282	
ID	ABB64282 standard; Protein; 1693 AA.
XX	XX
AC	ABB64282;
XX	XX
DT	25-MAR-2002 (first entry)
XX	XX
DE	Drosophila melanogaster polypeptide SEQ ID NO 19638.
XX	XX
KW	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical.
XX	XX
QS	Drosophila melanogaster.
XX	XX
PN	WO200171042-A2.
XX	XX
PD	27-SEP-2001.
XX	XX
PF	23-MAR-2001; 2001WO-US09231.
XX	XX
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.

11-OCT-2001.
30-MAR-2001: 2001WO-US08631.
31-MAR-2000: 2000US-0540217.
23-AUG-2000: 2000US-0649167.
(HYSE-) HYSEQ INC.
Drmanac RT, Liu C, Tang YT;
WPI: 2001-639362/73.
N-PSDB: AAS77392.
New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -
Claim 20; SEQ ID No 43564; 103pp; English.
The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.
Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published.pct.sequences.

```

Query Match      6.2%; Score 102; DB 22; Length 430;
Best Local Similarity 21.9%; Pred. No. 0.25;
Matches 61; Conservative 39; Mismatches 94; Indels 84; Gaps 14;

35  IRTETLASRGKPKSGKCOQKQKPTLRMGVAKLERQRIEERKKQLAAATVCDTSSVAS 94
   || : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
70  IRDKLMERNRRTG-----RTEKARIEWTDRTVRTWIGE--AVAA 108
   || : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
95  ISNNATRLPVDPGVVLQGFPPSLGSNRIYCGVGSGQVMIDPVTSPGFEVETSTTHE 154
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
109 AAADGVTSVPVPT--HTRHYAMHMLYAG-----IPLKVLQSLMGHKHSISST--- 155
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
155 LSSISNPMQMFNASSNNRCD--TCF-----KKKRLDQGNVNRSGNGGFS 197
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
156 -EYTKVPALDVAARHRCNDNLKTCHTSHGSMVAETAIVNHKKRKNPR--IVQSNDLTEA 212
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
198 KYTIMPPPMNGYDQ---YLLQSDHHQSGQGLFYDH-----RIARAASVASSTTINPY 247
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
213 AYSL-----SRDQKRMLYLF-VQDIRKSDGTLOEHDGICEIHVAKYAEIFG----- 257
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
248 FNEATNHTGPMEEFGSYMEGN--PRNGSGGVKYEYFFP 283
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
258 ---LTSAASTEEFERGEGSGNYRDEEDAGDEKGVESP 292
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

```

RESULT 5
ABG13382

351 SIELEADAAAEIITEDQVHEEVLSSNSTESSEVNVNAIA-----KAYAECC 400
130 GSGQVMIDPVSIP-----WGFEVTSSTTHLSSISNPQMFNASSNNRCDTC 175
401 TSSAAMLOPILQAAIKATPKVEVERIHEKIVKVKTEKHDLPSCOTAPHYQHSRQNETI 460
176 PKK-KRLDGDQNNVRSNGGFSKYTIMPPMNGYDQYLQSDHHQRS---QGFLYDHRI 231
461 VQAIRHLEGD--HLFADGSGGQKFG-----QOONLQOHQHHQMGVHHQLOHRL 509
232 AAA-----SVASSTTINP 245
510 AQDAPLALTITGQHAALAEPLQLKSGNGKQVITAKTAKDGGGLVSTSTSTAVTP 569
247 --YFNEAT-----NHTGPMEEFGSYMEGNPRNGSGVKEYEFPFGKYGERVSVWATTS 297
570 TAIKFKVQTGGHNGSHLPAGSSTGTITGS--NGSGAQO-----LTITTS 614
298 SLVGDGSPNTI 308
615 --LKQCRPGVI 623

RESULT 7
ABB71963
D ABB71963 standard; Protein; 1406 AA.
X C ABB71963;
X C
T 26-MAR-2002 (first entry)
E Drosophila melanogaster polypeptide SEQ ID NO 42681.
W Drosophila; developmental biology; cell signalling; insecticide;
W pharmaceutical.
S Drosophila melanogaster.
X WO200171042-A2.
X 27-SEP-2001.
X 23-MAR-2001; 2001WO-US09231.
X 23-MAR-2000; 2000US-191637P.
X 11-JUL-2000; 2000US-0614150.
X (PEKE) PE CORP NY.
X Venter JC, Adams M, Li PWD, Myers EW;
X WPI; 2001-656860/75.
X N-PSDB; ABL16066.
X New isolated nucleic acid detection reagent for detecting 1000 or more
X genes from Drosophila and for elucidating cell signalling and cell-cell
X interactions -
X Disclosure; SEQ ID NO 42681; 2lpp + Sequence Listing; English.
X The invention relates to an isolated nucleic acid detection reagent
X capable of detecting 1000 or more genes from Drosophila. The invention is
X useful in developmental biology and in elucidating cell signalling and
X cell-cell interactions in higher eukaryotes for the development of
X insecticides, therapeutics and pharmaceutical drugs. The invention
X discloses genomic DNA sequences (ABL16176-ABL30511); expressed DNA
X sequences (ABL01840-ABL16175) and the encoded proteins
X (ABB57737-ABB72072).
X The sequence data for this patent did not form part of the printed
X specification, but was obtained in electronic format directly from WIPO
X at ftp.wipo.int/pub/published_pct_sequences.
X Sequence 1406 AA;

Query Match 6.2%; Score 102; DB 22; Length 1406;
Best Local Similarity 23.0%; Pred. No. 1.4;
Matches 41; Conservative 28; Mismatches 63; Indels 46; Gaps 8;
QY 8 MSTQNSVGNPNDLLNRLVNVSSGEIRTEILKSGRKPGSKTGQOKQKPTLRGMVA 67
Db 358 VETKQVDTEFTPVLSN---LIKTEAPIKTES-----TEKPKPLVNGEAI 401
QY 68 KLERQRTEEKKOLAAATVGDTSVASISNATRLPVPDPGVVLOGFPSSLSGSRNIYCG 127
Db 402 KTR----DEVSRQL---NFGDKVS-----NSKLMPSPAP-----KSSSGSN----- 435
QY 128 GVSGQVMIDPVSIPNGFVETSTT-----HELSSISNPQMFNASSNNRCDTCFKKRL 181
Db 436 STASNHSSQSSSRKSYLSLSS 493

RESULT 8
ABB66224
ID ABB66224 standard; Protein; 729 AA.
X X ABB66224;
X X
X 26-MAR-2002 (first entry)
X Drosophila melanogaster polypeptide SEQ ID NO 25464.
X Drosophila; developmental biology; cell signalling; insecticide;
X pharmaceutical.
X Drosophila melanogaster.
X WO200171042-A2.
X 27-SEP-2001.
X 23-MAR-2001; 2001WO-US09231.
X 23-MAR-2000; 2000US-191637P.
X 11-JUL-2000; 2000US-0614150.
X (PEKE) PE CORP NY.
X Venter JC, Adams M, Li PWD, Myers EW;
X WPI; 2001-656860/75.
X N-PSDB; ABL10327.
X New isolated nucleic acid detection reagent for detecting 1000 or more
X genes from Drosophila and for elucidating cell signalling and cell-cell
X interactions -
X Disclosure; SEQ ID NO 25464; 2lpp + Sequence Listing; English.
X The invention relates to an isolated nucleic acid detection reagent
X capable of detecting 1000 or more genes from Drosophila. The invention is
X useful in developmental biology and in elucidating cell signalling and
X cell-cell interactions in higher eukaryotes for the development of
X insecticides, therapeutics and pharmaceutical drugs. The invention
X discloses genomic DNA sequences (ABL16176-ABL30511); expressed DNA
X sequences (ABL01840-ABL16175) and the encoded proteins
X (ABB57737-ABB72072).
X The sequence data for this patent did not form part of the printed
X specification, but was obtained in electronic format directly from WIPO
X at ftp.wipo.int/pub/published_pct_sequences.
X Sequence 729 AA;

Query Match 6.2%; Score 101.5; DB 22; Length 729;
Best Local Similarity 25.2%; Pred. No. 0.61;
Matches 72; Conservative 36; Mismatches 119; Indels 59; Gaps 13;

QY 6 FFMSTQNSVGN-PNDLLNRLVYNSGGEIRTTETLKSGRKPGSKTGQOKKPTLRGM 64
Db 112 FLISVNLVSKLPHEV--VOLIGNSGFIRMOIAENY-----YSDSDEENAHATLRGQ 165
QY 65 GV-----AKLERORIEEEKKOLAAATVGTSSVASISNNATRLPVPDPGV 110
Db 166 LLAASLRHKPFLHHKAKHLRNSPOKLNPPPEAVPHKSKSS-PDHPILKPVLEDP-- 222
QY 111 VLOGFPSSLSGSRNYICGGVGGQVMDPVISPWGFEVETSTTHELSSISNPMQFNASSN- 169
Db 223 -----PLTANLSKAADVANSAMVRAGS--AALECRVIVGLGTIEMPKQISHSKL 273
QY 170 NRCDTCEKFKRLDQNNVRSNGGFSKYTMIPPPMNGYDQYLLQSDHQRSGGFLYDH 229
Db 274 QTVRSCTIRKLRQKROPTIV-----LMCITP-----DSLQSS-----SGVLATY 315
QY 230 RIARAASVSASSTTINPYENATN--HTGPMEEFGSYMEGNPRNGS 273
Db 316 SSARLNFVSSSESENRFGLVTSVAVHTQIEE-----EYEPSAGS 356

RESULT 9
ABB71639
ID ABB71639 standard; Protein; 1059 AA.
XX AC ABB71639;
XX AC ABB71639;
DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 41709.
DE Drosophila;
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR N-PSDB; ABL15742.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -
XX Disclosure; SEQ ID NO 41709; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1059 AA;
XX Query Match 6.1%; Score 100; DB 22; Length 1059;

Best Local Similarity 17.5%; Pred. No. 1-5;
Matches 50; Conservative 45; Mismatches 120; Indels 70; Gaps 9;

QY 8 MSTDQNSVGNPNLLR-NTRLVYNSGGEIRTTETLKSGRKPGSKTGQOKKPTLRGMGV 66
Db 605 MEQQQHPRSNNSNRGSSSLAGTSGNNGKRNSTGSSSSSSGVETQQ----- 655
QY 67 AKLERORIEEEKKOLAAATVGTSSVASISNNATRLPVPDPGVVLOGFPSSLSGSRNYIC 126
Db 656 ---DRQR-----STTPMSSNRSSASSTT-----PPKLAASFVPAATSGS--- 692
QY 127 GGVSGQVMIDPVISPWGFEVETSTTHELSSISNPMQFNASSNRCDTCEKFKRL----- 181
Db 693 -ASGTSKORMPPOOSDYNYSNNAQHVASN-----SKRMSGVGP 732
QY 182 -DGDQNNVRSNGGFSKYTMIPPPMNGYDQYLLQSDHQRSGGFLYDHRIARAASVSAS 240
Db 733 SCSNGQQRQSASGNSGYQQVPPPSNSRSSSTQQQNHQKQKQ-----VQKQAPSOQ 785
QY 241 STTINPYFNATNHTGPMEEFG-----SYMEGNPRNGSGGVKEYEF 281
Db 786 QQQQQQQYHQCAKHPSPSQOLAAAAAHAYAHATADTSSATPRYDF 830

RESULT 10
ABB61226
ID ABB61226 standard; Protein; 607 AA.
XX AC ABB61226;
XX AC ABB61226;
DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 10470.
DE Drosophila;
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR N-PSDB; ABL05329.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -
XX Disclosure; SEQ ID NO 10470; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins
(ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
XX
XX

[illegible]

Mon Jan 27 14:05:55 2003

us-09-701-023-4.rag

Page 10

Db 1017 CQYKPDGSG 1026
: :|||

Search completed: January 26, 2003, 00:03:36
Job time : 78 secs

M protein - protein search, using sw model

Run on: January 26, 2003, 00:03:42 : Search time 23 Seconds

(without alignments)
401.687 Million cell updates/sec

Title: US-09-701-023-4

Perfect score: 1634

Sequence: 1 MATSLFFNSTDQNSVGNPND.....TTSSLVGDSPNTIDLSKL 314

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents,AA:*

1: /cgn2_5/ptodata/1/iaa/5A_COMB.pep.*

2: /cgn2_5/ptodata/1/iaa/5B_COMB.pep.*

3: /cgn2_5/ptodata/1/iaa/6A_COMB.pep.*

4: /cgn2_5/ptodata/1/iaa/6B_COMB.pep.*

5: /cgn2_5/ptodata/1/iaa/PCTUS_COMB.pep.*

6: /cgn2_5/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	93	5.7	656	2	US-08-343-443B-2
2	93	5.7	656	4	US-09-214-564A-4
3	92	5.6	1464	4	US-08-891-640-2
4	91	5.6	734	4	US-09-185-160-13
5	86.5	5.3	1250	1	US-08-441-139-9
6	86	5.3	632	1	US-08-443-104-4
7	86	5.3	632	1	US-08-442-859-4
8	86	5.3	632	2	US-08-398-489-4
9	86	5.3	632	5	PCT-US95-05534-4
10	85.5	5.2	349	4	US-09-149-476-371
11	85.5	5.2	422	4	US-09-071-035-484
12	85.5	5.2	449	4	US-09-071-035-482
13	85	5.2	654	1	US-08-083-590A-16
14	85	5.2	654	2	US-08-346-128-35
15	85	5.2	654	3	US-08-532-384-16
16	85	5.2	911	4	US-08-460-269C-4
17	85	5.2	2523	1	US-08-185-432-18
18	85	5.2	2523	4	US-08-899-232-3
19	83.5	5.1	899	4	US-08-470-335-249
20	83.5	5.1	899	4	US-08-467-602-301
21	83.5	5.1	933	4	US-08-467-602-343
22	83	5.1	594	3	US-08-729-416C-17
23	83	5.1	650	4	US-08-467-602-280
24	83	5.1	666	1	US-08-083-590A-17
25	83	5.1	666	2	US-08-346-128-36
26	83	5.1	666	3	US-08-532-384-17
27	83	5.1	1389	2	US-08-619-198-5

28 83 5.1 2703 1 US-08-185-432-19 Sequence 19, Appl
29 83 5.1 2703 4 US-08-899-232-4 Sequence 4, Appl
30 83 5.1 6095 4 US-09-144-085-2 Sequence 2, Appl
31 82.5 5.0 513 4 US-09-100-193-3 Sequence 3, Appl
32 82 5.0 852 4 US-08-470-335-248 Sequence 248, App
33 82 5.0 852 4 US-08-467-602-300 Sequence 300, App
34 82 5.0 886 4 US-08-467-602-342 Sequence 342, App
35 81.5 5.0 603 4 US-08-467-602-279 Sequence 279, App
36 81.5 5.0 657 4 US-09-355-166-1 Sequence 1, Appl
37 81 5.0 683 4 US-09-620-412C-357 Sequence 357, App
38 81 5.0 1297 2 US-08-290-731C-4 Sequence 4, Appl
39 81 5.0 1776 4 US-09-556-877-179 Sequence 179, App
40 81 5.0 1776 4 US-09-620-412C-179 Sequence 2, Appl
41 80.5 4.9 388 2 US-08-894-772-2 Sequence 2, Appl
42 80.5 4.9 388 2 US-09-207-844-2 Sequence 2, Appl
43 80.5 4.9 620 1 US-08-484-105-4 Sequence 4, Appl
44 80.5 4.9 620 1 US-08-484-106-4 Sequence 4, Appl
45 80.5 4.9 635 4 US-08-470-335-247 Sequence 247, App

ALIGNMENTS

RESULT 1

US-08-343-443B-2

; Sequence 2, Application US/08343443B

; Patent No. 5968734

; GENERAL INFORMATION:

; APPLICANT: Aurias, Alain

; APPLICANT: Delattre, Olivier

; APPLICANT: Desmaze, Chantal

; APPLICANT: Melot, Thomas

; APPLICANT: Peter, Martine

; APPLICANT: Ploougastel, Beatrice

; APPLICANT: Thomas, Gilles

; APPLICANT: Zucman, Jessica

; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF

; TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL

; TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS

; TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID

; NUMBER OF SEQUENCES: 129

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Weiser & Associates

; STREET: 230 South Fifteenth Street

; CITY: Philadelphia

; STATE: PA

; COUNTRY: USA

; ZIP: 19102

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: AEDIT 1.0 DOS text editor

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/343,443B

; FILING DATE: 18-NOV-1994

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/FR93/00494

; FILING DATE: 19-MAY-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: FR 92/06123

; FILING DATE: 20-MAY-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Weiser, Gerard J.

; REGISTRATION NUMBER: 19,763

; REFERENCE/DOCKET NUMBER: 989.6121P

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-875-8383

; TELEFAX: 215-875-8394

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

LENGTH: 656 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-343-443B-2

Query Match 5.7%; Score 93; DB 2; Length 656;
Best Local Similarity 20.2%; Pred. No. 0.5;
Matches 67; Conservative 34; Mismatches 127; Indels 104; Gaps 13;

QY 16 GNPNDLLRNLRLVNSSGEIRTELTKSRGKRP---GSKTQQQKQKPTLRGMGVAKLERQ 72
DB 45 GQPTDVSYTOAQTATYQ--TAVATSYGQPPPTGYTTPTAPQAYSQ-P-VQGYGTGAYD-- 99
QY 73 RIEEEKQLAAATVCDT-SSVASISNATRLPVPV-----DPGVVLOG 114
DB 100 -----TTATVTTTAAOASAYGTQPAYCQQAATAPTRPDGKPTTSQP 151
QY 115 FPSLSGSRNIYCGVGSQ-----VMIDPVISPWGPFVETSTTHELSSI 158
DB 152 QSSTGGYNO---PSLGGQSNYSYVPQVPSYPMQPTAPPSYPTSYSTQPTSYDOSSY 208
QY 159 SNPOMFNASSNNRCDTCFKKRLDQNNVVRNSGGGFSKYTMIPPPMNGY----DOYLL 214
DB 209 SQQNTYGOPSSY-----GQSSYGOQSSYGOQPTSYPPQTGYSQAPSYSQ 256
QY 215 QSDHHQSQGLFDHRIARAASVSASTTINPYNEATNHTGPM-----259
DB 257 QSSSYGOQSSYFRODH-----PSSMGVYGQESGGFSGFGENRSMGPDNRGRG 305
QY 260 --EFGSYMEGNPRNGSGVKEYEFPCKYGER 289
DB 306 GFDGRGMSRGRGGRGGM-----GSAGER 330

RESULT 2

US-09-214-564A-4
; Sequence 4, Application US/09214564A
; Patent No. 6150515
; GENERAL INFORMATION:
; APPLICANT: Sharp, Phillip A.
; APPLICANT: Zhou, Qiang
; TITLE OF INVENTION: Tat-SF: Cofactor For Stimulation Of Transcriptional
; FILE REFERENCE: Elongation By HIV-1 Tat
; CURRENT APPLICATION NUMBER: US/09/214,564A
; PRIOR FILING DATE: 1999-08-18
; PRIOR FILING DATE: 1996-07-03
; PRIOR APPLICATION NUMBER: US 60/021,218
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: PCT/US97/11713
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 656
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-564A-4

Query Match 5.7%; Score 93; DB 4; Length 656;
Best Local Similarity 20.2%; Pred. No. 0.5;
Matches 67; Conservative 34; Mismatches 127; Indels 104; Gaps 13;

QY 16 GNPNDLLRNLRLVNSSGEIRTELTKSRGKRP---GSKTQQQKQKPTLRGMGVAKLERQ 72
DB 45 GQPTDVSYTOAQTATYQ--TAVATSYGQPPPTGYTTPTAPQAYSQ-P-VQGYGTGAYD-- 99
QY 73 RIEEEKQLAAATVCDT-SSVASISNATRLPVPV-----DPGVVLOG 114

QY 115 FPSLSGSRNIYCGVGSQ-----VMIDPVISPWGPFVETSTTHELSSI 158
DB 152 QSSTGGYNO---PSLGGQSNYSYVPQVPSYPMQPTAPPSYPTSYSTQPTSYDOSSY 208
QY 159 SNPOMFNASSNNRCDTCFKKRLDQNNVVRNSGGGFSKYTMIPPPMNGY----DOYLL 214
DB 209 SQQNTYGOPSSY-----GQSSYGOQSSYGOQPTSYPPQTGYSQAPSYSQ 256
QY 215 QSDHHQSQGLFDHRIARAASVSASTTINPYNEATNHTGPM-----259
DB 257 QSSSYGOQSSYFRODH-----PSSMGVYGQESGGFSGFGENRSMGPDNRGRG 305
QY 260 --EFGSYMEGNPRNGSGVKEYEFPCKYGER 289
DB 306 GFDGRGMSRGRGGRGGM-----GSAGER 330

RESULT 3

US-08-891-640-2
; Sequence 2, Application US/08891640
; Patent No. 6288173
; GENERAL INFORMATION:
; APPLICANT: Chambon, Pierre
; APPLICANT: Gronemeyer, Hinrich
; APPLICANT: Voegel, Johannes
; APPLICANT: Lutz, Yves
; TITLE OF INVENTION: Transcriptional Intermediary Factor-2
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/891,640
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/021,247
; FILING DATE: 12-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1383.0130001/EKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1464 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-891-640-2

Query Match 5.6%; Score 92; DB 4; Length 1464;
Best Local Similarity 17.9%; Pred. No. 2.2;
Matches 56; Conservative 45; Mismatches 107; Indels 104; Gaps 10;

QY 13 NSVGNPNLLR-----NTRLVVNSSGEIRTELTKSRGKRPQSKTQQQKQKPTLRG 63
DB 1211 SNVSNVNLRLPCVPTQAPINAQOROEILNOHLRORQ---HQOQVOORTLNRG 1267
QY 64 MGVAKLERQRIEEEKQLAAATVCDTSSVASISN-----NATRLPVPVDPGVVLOGPFS 117

Y 118 SLGSRNYCGVGSGQVMDPVPISPKGFVETSTTHLSSINPQMFNASSNNRCDTCFK 177
b 1315 FTGAT-----TPQSPLMSP-----RMAHTQSPMMQSQAN-----P 1345
Y 178 KKRLDQNNVRSNGGSGFKYMTIPPPMNGYDQYLQSDHQRSGFLYDHRITARAASV 237
b 1346 AYQAPSDINGWAGNMGNSWFSQSPPHFG-----QANT 1381
Y 238 SASSTTINPNEATNHTGMEERFSGYMEGNPRNGSGVKEYEYFFFGKGYGERSVWATTS 297
b 1382 SMYSNNNNINVSMAFN-----TGMSMNMQMTGQI-SMTSVTSVST 1421
Y 298 SLVGDCSPNTID 309
b 1422 SGLSSMGPEQVN 1433

RESULT 4

US-09-185-160-13

Sequence 13, Application US/09185160

Patent No. 6252137

GENERAL INFORMATION:

APPLICANT: ODELL, JOAN T.

APPLICANT: HARDER, PATRICIA A.

TITLE OF INVENTION: SOYBEAN HOMOLOG OF A SEED-SPECIFIC

TITLE OF INVENTION: TRANSCRIPTION ACTIVATOR FROM PHASEOLUS

TITLE OF INVENTION: VULGARIS

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY

STREET: 1007 MARKET STREET

CITY: WILMINGTON

STATE: DELAWARE

COUNTRY: USA

ZIP: 19898

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.50 INCH

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: MICROSOFT WINDOWS 95

SOFTWARE: MICROSOFT WORD VERSION 7.0A

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/185,160

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/065,459

FILING DATE: NOVEMBER 12, 1997

ATTORNEY/AGENT INFORMATION:

NAME: MAJARIAN, WILLIAM R.

REGISTRATION NUMBER: 41,173

REFERENCE/DOCKET NUMBER: BB-1096

TELECOMMUNICATION INFORMATION:

TELEPHONE: 302-992-4926

TELEFAX: 302-773-0164

INFORMATION FOR SEQ ID NO: 13:

LENGTH: 734 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-185-160-13

Query Match 5.6%; Score 91; DB 4; Length 734;

Best Local Similarity 21.8%; Pred. No. 0.97;

Matches 70; Conservative 38; Mismatches 83; Indels 130; Gaps 16;

Y 5 LFFMSTQNSVGNPNLLRNLRLVNVSSGEIRTTTLKSRGRK-PSGKTGQOKKQKPLRG 63

b 228 LEWLKSNKDSV-SAND-LRNVL-----KKATLESAAARRLGGREAMKQLKLEW 277

Y 64 MGVAKLERQRIEEKKQLAAATVGTDTSSVASISNNATRLPVPVDPGVVLOG-FPSSLSGN 122

Db 278 VQTSHLONKRKE-----NNGSSISS-----VLAQFQDPGQCN 311
QY 123 RIYCGVGSGQVMDPVPISPKGFVETSTTHLSSINPQMFNASSNNRCDTCFKKRLD 182
Db 312 -----NNQNTQSGSPAPESNTCF----- 329
QY 183 GDQNNVRSNGGSGFKYMTIPPPMNGYDQYLQSDHQRSG-----GFLYDHRITARAASVSA 239
Db 330 --NQTPRLSSQTFAT-----DQALLMVPQQFPQPMVGVGDPYTSGAASNNI 376
QY 240 SST-----TINPY-----FNEATNHTGPMEEFGSYMEGNPRNGSGG 275
Db 377 SATNNHNSNPYQGAQGYHMLSAHWSPHSLFNVASNYS---CSFGDNNGLNPHGGFGG 433
QY 276 -----VKEYEFPFGKYGERV 290
Db 434 GGYGNHNPYQFPFGP-GDRL 453

RESULT 5

US-08-441-139-9

Sequence 9, Application US/08441139

Patent No. 5773245

GENERAL INFORMATION:

APPLICANT: Wittup, Dr. Karl D.

APPLICANT: Robinson, Anne S.

TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF

TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER

STREET: 400 Garden City Plaza

CITY: Garden City

STATE: NY

COUNTRY: USA

ZIP: 11530

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/441,139

FILING DATE: 15-MAY-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/089,997

FILING DATE: 06-JUL-1993

ATTORNEY/AGENT INFORMATION:

NAME: Digiglio, Frank S.

REGISTRATION NUMBER: 31,346

REFERENCE/DOCKET NUMBER: 8646

TELECOMMUNICATION INFORMATION:

TELEPHONE: 516-742-4343

TELEFAX: 516-742-4366

TELEX: 230 901 SANS UR

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 1250 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-441-139-9

Query Match 5.3%; Score 86.5; DB 1; Length 1250;

Best Local Similarity 19.9%; Pred. No. 6.7;

Matches 81; Conservative 49; Mismatches 138; Indels 139; Gaps 17;

QY 2 ATSLFFMSTQNSVGNPNLL-----RNLRLVNVSSGEIRTTTLKSRGRKPSGKTGQOKK 58

Db 453 ATSNFLSSPSSDLSKDDLVSVERKRSSTINDDSLSSPTKSGVRRRSLKORPTQKK 512

```

Qy 59 ----PTURGMGVAKLERQRIEEKK-----QLAAATVG---DTSSVASIS 96
Db 513 NDDVEVEGSSLIVEEIENDKYKPLYAGHVAVLDRTGPQSGTGLLRPSQOASDN 572
Qy 97 NNATRPL----VPDPGVVLOGFPSSLG-----SNRIYCGGVGSQVMIDPVI 140
Db 573 NKPPQSPQIAWFKEPTDKKYPLIAITELAPKDFVENADKIYSEKLEVASIKRW----PII 627
Qy 141 S---PWGFVFETSTTHELSSISNPQFNASSNNRCDTFCKKRKLDCDQNNVRNSGGGFSK 198
Db 628 SLHPPFGIL----VSELGIDHP-----DTEIDSILRDN--NFLS 660
Qy 199 YTIMPPMNGDYOLYLQSDHQRSGFLYDHRIAASVSASSTINPYFNEATNHTGPM 258
Db 661 -----NEYLDQKNPQEKFSFQLPPLPATBSLEYRRNFTOTNEYINFALSELGVW 709
Qy 259 EEF-----GYMECPNRNGSGV----- 276
Db 710 SEFALHVRNNGNTLEGCHVVVDVTSHIEEGSSVDRRAKRKRSAVFMPOKLVNLLPQSFN 769
Qy 277 KEYEFPFGYKERYSVAT-----TSSUGD--CSPNTIDLSLK 313
Db 770 DELSLAPGKEAFSLVYVYTLDSSTLRKTGWGESTIPSNLT-LSLE 815

```

RESULT 6

```

US-08-443-104-4
; Sequence 4, Application US/08443104
; Patent No. 5691162
; GENERAL INFORMATION:
; APPLICANT: Shuster, Jeffrey R.
; APPLICANT: Madden, Mark
; APPLICANT: Moyer, Donna L.
; APPLICANT: Fuglsang, Claus
; APPLICANT: Branner, Sven
; TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
; TITLE OF INVENTION: ACTIVITY
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 5691162o NO. 5691162disk of No. 5691162th America, Inc.
; STREET: 405 Lexington Avenue, 64th floor
;

```

RESULT 7

RESOLUTION 1
US-08-442-859-4
Sequence 4, Application US/08442859
Patent No. 5807729
GENERAL INFORMATION:
APPLICANT: Shuster, Jeffrey R.
APPLICANT: Madden, Mark
APPLICANT: Moyer, Donna L.
APPLICANT: Fuglsang, Claus
APPLICANT: Branner, Sven
TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
TITLE OF INVENTION: ACTIVITY
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5807729Q No. 5807729disk of No. 5807729th America, Inc.
STREET: 405 Lexington Avenue. 64th Floor

Query Match	5.3%	Score 86;	DB 1;	Length 632;
Best Local Similarity	22.9%;	Pred. No. 2.6;		
Matches	61;	Conservative 39;	Mismatches 104;	Indels 62; Gaps
QY	44	GRPGSKTGOOK-----KPTLRGMGVAKLEPQRIEEKKOLAAATVGDTSVASISNN	98	
	:	:: :: :	: : : :	
	:	: : : :	: : : :	
Db	138	GKIPGLTKRDEKDPDALKDTVDVLSPVEADKAKEKKSNHYTF--TGKTGVTSPK	194	
QY	99	ATRLPVYPDGVQLQ---GFPSLSGNRI---YCGVGSGOV---MIDPV-----ISPWF	145	
	:	: : : :	: : : :	
	:	: : : :	: : : :	
Db	195	EAKLTYLVDENKELKLTWRVETDIVDNLLTYNAAKTDVGVGVDAEATKYVPWG	254	
QY	146	VETSSTTHELSSITSNQMFNAS-----SNNRCDTCFKKKLDDQNMYVRNSGGGS	197	
	:	: : : :	: : : :	
	:	: : : :	: : : :	
Db	255	NDFSKGR--STVENFWNLAASEFTWLSGNSNYTTT-----RGNGIAQVNPSGGS	304	
QY	198	KYTMIPPPMMGYDQYLQSDHHHRSOGFLYDHRIA-----RAASVSASSITTINPFNE	250	
	:	: : : :	: : : :	
	:	: : : :	: : : :	
Db	305	TY-----LNRYRP-----DSPSLKFEDYSTSTITPTTYRDASIAQLFYTANKY-HD	350	
QY	251	ATNHTGPMEFEGSYMEGNPNRNGSGV	276	
	:	: : : :	: : : :	
	:	: : : :	: : : :	
Db	351	LLYLLGFTEOAGNFOTNN--NGOGV	374	

```

QY 44 GRKPGSKTGQKQK-----KPTLRGMGVAKLERQRIEKKKOLAATVGDTSVASISNN 98
   || || || || || || || || || || || || || || || || || || || || ||
Db 138 GKTPGLPLTRDEKDPDALKDTVDVLUL--PVEADRAKAEEKSKNNHYTF--TGTKGTVPKP 194
   || || || || || || || || || || || || || || || || || || || || ||
QY 99 ATRLPVPDGPVVLO---GPPSSLSGNRI--YCGGVGSGGV--MIDPV-----ISPWGF 145
   || || || || || || || || || || || || || || || || || || || || ||
Db 195 EAKTLVLDENKELKUTWRVEFDIVDNWLLTTVNAAKTDEVGVVDVYVNEATYKVTPMGV 254
   || || || || || || || || || || || || || || || || || || || || ||
QY 146 VETSSSTHELSSISNPOMFNAS-----SNNKCDTCFKKKRLDQNNVRSNGGGFS 197
   || || || || || || || || || || || || || || || || || || || || ||
Db 255 NDRPSKSR--STVENPWNLAASEFTWLSPGSNNYTT-----RGNNGIAQVNPSSGS 304
   || || || || || || || || || || || || || || || || || || || || ||
QY 198 KYTMIPPPNGYDQYLQSDHHQRSOGFELYDHRIA-----RAASVSASSTTINPYFNE 250
   || || || || || || || || || || || || || || || || || || || || ||
Db 305 TY-----LNNYRP-----DPSPLKFEYDYSTSTPTTPTTYRDASIAIOLFTYANKY-HD 350
   || || || || || || || || || || || || || || || || || || || || ||
QY 251 ATNHTGPMEEFGSYMEGNPRNGSGV 276
   || || || || || || || || || || || || || || || || || || || || ||
Db 351 LLYLLGTFQAGNFQTN--NGOGGV 374
   || || || || || || || || || || || || || || || || || || || || ||

RESULT 9
PCT-US95-05534-4
: Sequence 4, Application PC/TUS9505534
: GENERAL INFORMATION:
: APPLICANT:
: APPLICANT:
: TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
: TITLE OF INVENTION: ACTIVITY
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Novo Nordisk of North America, Inc.
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Tape
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/05534
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/238/108
: FILING DATE: 04-MAY-1994
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/398,489
: FILING DATE: 03-MARCH-1995
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Agtis Dr., Cheryl H.
: REGISTRATION NUMBER: 34,086
: REFERENCE/DOCKET NUMBER: 4180.204-WO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 632 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
PCT-US95-05534-4

Query Match 5.3%; Score 86; DB 5; Length 632;
Best Local Similarity 22.9%; Pred. No. 2.6;

```


[illegible]

Db 186 SKAGLEKQAAAEQARLAAEQAAAEKAAKQAAKAPKAEVKAEPVASSSTTEAQA 245
QY 105 PVDPGVLOGFPSSLSGRIYCGVGSGQWMDPVISPWGCVETSTTHLSISNPQMF 164
Db 246 -----SSSSA-----
QY 165 NASSNNRCDTCFK---KKRLDGDNNVRSNGGFSKYTMIPPPMNGYDQYLLOSDHQR 221
Db 268 NSATENTGSSSEQVPQPTTPSDNGNNGGTGGGTVPPTPEPTPAPSADPTINALNLVLRQ 327
QY 222 SQGF-----LYDHRIARAASVSASSTTINPYFNEATNHTGPMEEFGS-----YMG 267
Db 328 SGLRPVVDAGLAASATARAQVEAGGIP-NDHWSRGDEVIAIMWAPGNSVIMAWYNET 386
QY 268 NPRNGSG-GVKEYPFGPKYGERVSVVATSSLVG 301
Db 387 NMVTASGSGHRDWEINPGI--TRVGFYSGSTIVG 419

RESULT 12

US-09-071-035-482
; Sequence 482, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: GIL H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 482:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 449 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-09-071-035-482

Query Match 5.2%; Score 85.5; DB 4; Length 449;

Best Local Similarity 22.5%; Pred. No. 1.7;

Matches 62; Conservative 31; Mismatches 119; Indels 63; Gaps 9;

QY 49 SXTGQOKKQKPTLRGMGVAKLERQRIEKKQLAA---ATVCDTSSVASISNNATRLPV 104
Db 213 SKAGLEKQAAAEQARLAAEQAAAEKAAKQAAKAPKAEVKAEPVASSSTTEAQA 272
QY 105 PVDPGVLOGFPSSLSGRIYCGVGSGQWMDPVISPWGCVETSTTHLSISNPQMF 164
Db 273 -----SSSSA-----

QY 165 NASSNNRCDTCFK---KKRLDGDNNVRSNGGFSKYTMIPPPMNGYDQYLLOSDHQR 221
Db 295 NSATENTGSSSEQVPQPTTPSDNGNNGGTGGGTVPPTPEPTPAPSADPTINALNLVLRQ 354
QY 222 SQGF-----LYDHRIARAASVSASSTTINPYFNEATNHTGPMEEFGS-----YMEG 267
Db 355 SGLRPVVDAGLAASATARAQVEAGGIP-NDHWSRGDEVIAIMWAPGNSVIMAWYNET 413
QY 268 NPRNGSG-GVKEYPFGPKYGERVSVVATSSLVG 301
Db 414 NMVTASGSGHRDWEINPGI--TRVGFYSGSTIVG 446

RESULT 13

US-08-083-590A-16
; Sequence 16, Application US/08083590A
; Patent No. 5786158
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
; TITLE OF INVENTION: Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 654 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide

US-08-083-590A-16

Query Match

Best Local Similarity 18.6%; Pred. No. 3.5;

Matches 50; Conservative 40; Mismatches 103; Indels 76; Gaps 9;

QY 46 KPGSKTGQKKKPTLRGMGVAKLERQRIEKKQLAAATVCDTSSVASIS----- 96
Db 290 KPSVGS--KKARKPSIKNGCKEAKELKARRKKSGDGTLLDSSGSGVLSPVDSLESTH 347
QY 97 ---NNATRLPVDPVDPGVVLOGFP-----SSLSGSRNIYCG-----GVSGGVWIDPV 139
Db 348 GYLSVSVSPPLMTSFPQSPSPMLNHLTSMPSQLGMHINMATKQEMAAGSNRRMADAM 407
QY 140 ISPWGCVETSTTHLSISNPQMFNAS-----SNRCDTCFK-----KKRLDG 183
Db 408 V-----PRLTHLNASSPNTIMSGMHHFTVGVGAPTMSQCDMLARLONGWVQNOYDP 459

Y 184 DONNVVRNSG-----GGFSKYTMIPPPMNGYDQYLLOSDHHQ 220
b 460 IRNGIOGNAQAALQHLGMLTSLHNLGTPATLTSQMMTYQAMPNTRLANQPHLMQAQQMQ 519
Y 221 RSGGF-LYDHRITARAASVSASSTTINPYF 248
b 520 QQONLQHQSMQQHNSSTTSHINSPPF 548
RESULT 14
S-08-346-128-35
Sequence 35, Application US/08346128
Patent No. 5856441
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon et al.
TITLE OF INVENTION: Human No. 5856441ch And Delta, Binding Domains
TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/346,128
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/879,038
FILING DATE: 30-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-009
TELEPHONE: 212 8698864/9741
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 654 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
S-08-346-128-35
Query Match 5.2%; Score 85; DB 2; Length 654;
Best Local Similarity 18.6%; Pred. No. 3.5;
Matches 50; Conservative 40; Mismatches 103; Indels 76; Gaps 9;
Y 46 KPGSKTGQOKKPTLRGNGVAKLERQRIEKKQLAAATVGDTSVASIS-----96
b 290 KPSVQS--KKARKPSIKNGCKEAKELKARKKSDGKTTLLDSSGSGVLSPVDSLESTH 347
Y 97 ---NNATRLPVDPGVVLGGFP-----SSLGSNRIYCG-----GVGSGQVMIDPV 139
b 348 GYLSDYSSPPLMTSPFQQSPSPMLNLTSMPESQLGMNHINMATKQMAAGSNRMADFAM 407
Y 140 ISPWGPFVETSSSTHELSSISNPFMFNAS-----SNNRCDTCFK-----KKRLDG 183
b 408 V-----PRLTHLNASSPNTIMNSGHHFTVGGAPTWNISQCDWLARLQNGVMQNOYDP 459
Y 184 DONNVVRNSG-----GGFSKYTMIPPPMNGYDQYLLOSDHHQ 220
b 460 IRNGIOGNAQAALQHLGMLTSLHNLGTPATLTSQMMTYQAMPNTRLANQPHLMQAQQMQ 519

QY 221 RSGGF-LYDHRITARAASVSASSTTINPYF 248
Db 520 QQONLQHQSMQQHNSSTTSHINSPPF 548
RESULT 15
US-08-532-384-16
Sequence 16, Application US/08532384
Patent No. 6083904
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, S. et al.
TITLE OF INVENTION: Therapeutic And Diagnostic Methods
TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
TITLE OF INVENTION: Nucleic Acids
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/532,384
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/083,590
FILING DATE: 25-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-015
TELEPHONE: 212 8698864/9741
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 654 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-532-384-16
Query Match 5.2%; Score 85; DB 3; Length 654;
Best Local Similarity 18.6%; Pred. No. 3.5;
Matches 50; Conservative 40; Mismatches 103; Indels 76; Gaps 9;
QY 46 KPGSKTGQOKKPTLRGNGVAKLERQRIEKKQLAAATVGDTSVASIS-----96
Db 290 KPSVQS--KKARKPSIKNGCKEAKELKARKKSDGKTTLLDSSGSGVLSPVDSLESTH 347
QY 97 ---NNATRLPVDPGVVLGGFP-----SSLGSNRIYCG-----GVGSGQVMIDPV 139
Db 348 GYLSDYSSPPLMTSPFQQSPSPMLNLTSMPESQLGMNHINMATKQMAAGSNRMADFAM 407
QY 140 ISPWGPFVETSSSTHELSSISNPFMFNAS-----SNNRCDTCFK-----KKRLDG 183
Db 408 V-----PRLTHLNASSPNTIMNSGHHFTVGGAPTWNISQCDWLARLQNGVMQNOYDP 459
QY 184 DONNVVRNSG-----GGFSKYTMIPPPMNGYDQYLLOSDHHQ 220
Db 460 IRNGIOGNAQAALQHLGMLTSLHNLGTPATLTSQMMTYQAMPNTRLANQPHLMQAQQMQ 519
QY 221 RSGGF-LYDHRITARAASVSASSTTINPYF 248

Db 520 000NLQJHOSMOOOHHNSSTTSTHNSPF 548

Search completed: January 26, 2003, 00:07:17
Job time : 28 secs

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	99.5	6.1	676	10	US-09-801-368-302	Sequence 302, App
2	95	5.8	625	10	US-09-801-368-294	Sequence 294, App
3	92.5	5.7	1239	12	US-10-007-805-577	Sequence 577, App
4	90	5.6	1464	10	US-09-843-236-2	Sequence 2, Appli
5	92	5.5	742	10	US-09-732-180-9	Sequence 9, Appli
6	88.5	5.4	517	10	US-09-801-368-100	Sequence 100, App
7	86.5	5.3	660	12	US-10-007-693-139	Sequence 139, App
8	86.5	5.3	1250	10	US-09-801-368-364	Sequence 364, App
9	85.5	5.2	333	10	US-09-888-635-64	Sequence 64, Appl
10	85	5.2	941	9	US-09-855-754-4	Sequence 4, Appli
11	84.5	5.2	1325	10	US-09-741-659-304	Sequence 304, App
12	84.5	5.2	2383	10	US-09-912-020-302	Sequence 302, App
13	83.5	5.1	1016	10	US-09-815-242-5845	Sequence 5845, Ap
14	83	5.1	488	10	US-09-925-301-1409	Sequence 1409, Ap
15	82.5	5.0	382	10	US-09-801-368-210	Sequence 210, App
16	82.5	5.0	559	10	US-09-820-721A-1	Sequence 1, Appli
17	82.5	5.0	692	10	US-09-801-368-274	Sequence 274, App
18	82.5	5.0	1518	10	US-09-801-368-152	Sequence 152, App
19	82.5	5.0	1703	10	US-09-801-368-340	Sequence 340, App

QY 97 NNATRLPVPDVGVLQGFPS-SLGSNRIYGGVSGVMIDPVISPWGF--VETSTTH 153
Db 526 M-----FPDWSGPKSAKPGTNT-----NFGTFPPVOTAVNNG 558
QY 154 ELSSISNPOMFNASSNNRCDTCFKKRLDQNNVVRNSGGGFSKYTIMPPMNG--YDQ 211
Db 559 NSSNISSTNTNNNNNNNNNSNNGNDNNNSNNSYSNNE--DAPVNGAISE 616
QY 212 YLLQSDHHRQSGELYDHRIRARAASVSASSTTINYPNEATNHTG 256
Db 617 HTDGDNSNQNSSTYD-----AAATAYNGTGLTPYINTAQTPLG 657

RESULT 2

US-09-801-368-294
; Sequence 294, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 294
; LENGTH: 625
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-294

Query Match 5.8%; Score 95; DB 10; Length 625;
Best Local Similarity 21.4%; Pred. No. 0.52;
Matches 72; Conservative 39; Mismatches 139; Indels 86; Gaps 15;
QY 6 FPMSTDONSVCNPNDDLNTLRLVNSSGEIRTETLKSGRKPGSKTGQKQKPTLRGMG 65
Db 212 FGCSTCSKFKRPQDLKHLKHLESGLK-----RRGPKWGSKRTSKNK----- 259
QY 66 VAKLERQRIEKKQLAAATVGDTSVASISNNATRLPVPDVGVLQGFPSLSGNRIY 125
Db 260 --SCASDAVSSCSVSFAIAGSKFHSSTPQILPLPVGIS-----OHLPSQQQRAIS 312
QY 126 CGGVSGQV-MIDPVISPWGFVETSTTHELSSISNPQMFNASS-----NNRC-----D 173
Db 313 LNQLCSDLSQYKVPYSP-----QLSARLOTILPLYNNGSTVSQGANRSRMVYED 365
QY 174 TCFKKKRLDQDQ-----NNVRSNNGGFSKYT-----MIDPPMNGYDQYLQSDHH 219
Db 366 GCSNKTIANATQFTKLSRNMINNYILQQSGGSTSSSSGRIPVAQTSYVQ-PPNAPSY 424
QY 220 QRSQGLYDHRIRARAASVSASSTIN-----PYFNEATNHTGPMEEFGSYMEGNPR 270
Db 425 QSVGG-----GSSISATATATVPVRLAKYPTGPSLTELHPLHS----- 465
QY 271 NGSQGV--KEYEFPFGYGERVSVVATTSSLVGDSCS 304

Db 466 NTAGGVENRQSQVAMPHY---PSVRAAPSYSSSCGS 498

RESULT 3

US-10-007-805-577
; Sequence 577, Application US/10007805
; Patent No. US20020150581A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 577
; LENGTH: 1239
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-007-805-577

Query Match 5.7%; Score 92.5; DB 12; Length 1239;
Best Local Similarity 21.4%; Pred. No. 2.5;
Matches 64; Conservative 46; Mismatches 128; Indels 61; Gaps 14;
QY 19 NDLLRNLRLVNSSGEIRTETLKSGRKPGSKTGQKQKPTLRG---MGVAKLERQR-- 73
Db 957 NKYPEDIKILQEKNAELQM-TLKLKQKTVTKRASQYREQLKVLTAENTMLTSKLKEKQDK 1015
QY 74 -----IEEEKKOLAAATVGDTSVASISNNATRLPVPDVGVLQGFPSLSGNRIYCG 127
Db 1016 EILETEIESHPRLASALQDHDQSVTSRKQELAFHSAGD--APLOGIMNVDSNTIY-- 1071
QY 128 GVGSGGVNIDPVISPWGFVETSTTHELSSISNPQMFNA-----SSNNRCDT-CFKKK-- 179
Db 1072 ---NNEVLHQPLYEAQR--KSKSPKINLNTAGDDLRENALVSEHAQRDCTOCQMKAE 1126
QY 180 -RLDGDQNNVVRNSGGGFSKYTIMPPMNGYDQYLQSDHHR--SQGLYDHRIRARAAS 236
Db 1127 HMYQNEQDNV-----DKHT---EQQESLEQKLFQLESKNRWLRQOLVYAHK-----K 1170
QY 237 VSASSTTINYPF-----NEATNHTGPMEEFGSYMEGNPRNGSGGVKEYEFF 282
Db 1171 VNKSQVNTIQFPEMKMQRLHKEKNEVFYNGNHLKERIDQYEKAEAREVSIKKYKF 1229
RESULT 4
US-09-842-256-2
; Sequence 2, Application US/09842256
; Patent No. US20020106727A1
; GENERAL INFORMATION:
; APPLICANT: Chambon, Pierre
; Groenemeyer, Hinrich
; Voegel, Johannes
; Lutz, Yves
; TITLE OF INVENTION: Transcriptional Intermediary Factor-2
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW, Suite 600
; CITY: Washington
; STATE: DC

COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/842,256
FILING DATE: 26-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/891,640
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1383.0130001/EKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1464 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
S-09-842-256-2
Query Match 5.6%; Score 92; DB 10; Length 1464;
Best Local Similarity 17.9%; Pred. No. 3.5;
Matches 56; Conservative 45; Mismatches 107; Indels 104; Gaps 10;
Y 13 NSVGNPNDLR-----NTRLVNVSSGEIRTEITLKRGRKPGSGTGOKKKPTLRG 63
b 1211 SNVSNVNLTRPGVPTQAPINAOIAQORIEILNQHRLRQRM---HQOQGVQORTLMRG 1267
Y 64 MGAKLRQRRIEKKKOLAAATVGDTSVASISN-----NATRLPVPDPGVVYLGFP 117
b 1268 QGL-----NNTFSMVAPSNMPTMSNPRIQNAQOFPFPYVIGISQQPDGP 1314
Y 118 SLASNRYCGVSGGVQVMDPVISWPGFVETSTTHELSSINPQFNASSNRCDCFCF 177
b 1315 FTGAT-----TPQSLMSP-----RMAHTQSPMQSQAN-----P 1345
Y 178 KKRIDGDNVNVRSNGGFSKYTMIPPPMNGYDYLQSDHQRSGFLYDHRIRAAV 237
b 1346 AYQAPSDINGWAQNGMGSNMFSSQSPPHFG-----QQANT 1381
Y 238 SASSTTINPYENATNHTGPMEEFGSYMEGNPRNGSCGVKEYEFPFGKYGRVSVWATTS 297
b 1382 SWYNNMNNINYSMATN-----TGMSSMNQMTGQI-SMTSVTSVST 1421
Y 298 SLVGDCSPNTID 309
b 1422 SGLSSMGPEQVN 1433
RESULT 5
IS-09-732-180-9
Sequence 9, Application US/09732180
Patent No. US20020137702A1
GENERAL INFORMATION:
APPLICANT: Arkowitz, Robert A
APPLICANT: Nern, Peter MA
TITLE OF INVENTION: Yeast receptor
FILE REFERENCE: DYO025.001AUS
CURRENT APPLICATION NUMBER: US/09/732,180
CURRENT FILING DATE: 2000-12-07
PRIOR APPLICATION NUMBER: US 60/169,699
PRIOR FILING DATE: 1999-12-07
NUMBER OF SEQ ID NOS: 11

COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/842,256
FILING DATE: 26-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/891,640
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1383.0130001/EKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1464 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
S-09-842-256-2
Query Match 5.6%; Score 92; DB 10; Length 1464;
Best Local Similarity 17.9%; Pred. No. 3.5;
Matches 56; Conservative 45; Mismatches 107; Indels 104; Gaps 10;
Y 13 NSVGNPNDLR-----NTRLVNVSSGEIRTEITLKRGRKPGSGTGOKKKPTLRG 63
b 1211 SNVSNVNLTRPGVPTQAPINAOIAQORIEILNQHRLRQRM---HQOQGVQORTLMRG 1267
Y 64 MGAKLRQRRIEKKKOLAAATVGDTSVASISN-----NATRLPVPDPGVVYLGFP 117
b 1268 QGL-----NNTFSMVAPSNMPTMSNPRIQNAQOFPFPYVIGISQQPDGP 1314
Y 118 SLASNRYCGVSGGVQVMDPVISWPGFVETSTTHELSSINPQFNASSNRCDCFCF 177
b 1315 FTGAT-----TPQSLMSP-----RMAHTQSPMQSQAN-----P 1345
Y 178 KKRIDGDNVNVRSNGGFSKYTMIPPPMNGYDYLQSDHQRSGFLYDHRIRAAV 237
b 1346 AYQAPSDINGWAQNGMGSNMFSSQSPPHFG-----QQANT 1381
Y 238 SASSTTINPYENATNHTGPMEEFGSYMEGNPRNGSCGVKEYEFPFGKYGRVSVWATTS 297
b 1382 SWYNNMNNINYSMATN-----TGMSSMNQMTGQI-SMTSVTSVST 1421
Y 298 SLVGDCSPNTID 309
b 1422 SGLSSMGPEQVN 1433
RESULT 5
IS-09-732-180-9
Sequence 9, Application US/09732180
Patent No. US20020137702A1
GENERAL INFORMATION:
APPLICANT: Arkowitz, Robert A
APPLICANT: Nern, Peter MA
TITLE OF INVENTION: Yeast receptor
FILE REFERENCE: DYO025.001AUS
CURRENT APPLICATION NUMBER: US/09/732,180
CURRENT FILING DATE: 2000-12-07
PRIOR APPLICATION NUMBER: US 60/169,699
PRIOR FILING DATE: 1999-12-07
NUMBER OF SEQ ID NOS: 11

QY 111 -----VLQGF-----SSLGSRNYCGVGGQVNI-DPVISPW----- 143
Db 152 KEQGLKNIPLEWGEQTEIAANDPNRNLKSSRLP-AEHKGEAFIGSPVTSPOOMPTS 209
QY 144 -GFVETSTHESSTINQMFNASSNRCDTCFKKRLDGDQNNVRSNGGFGSKYIMI 202
Db 210 NPFRTSMFKEELESPPSDHYNESIFSQVDEPIKK-----EPNSAQFSKGIFS 259
QY 203 P-----PMNGYDOYLLQSDHH 219
Db 260 PQQOQLOOQOQMI-NQPSLNTSRDIESDSNKECLNSGVSPTSNRFSLSFSDSN 319
QY 220 QR-----SQGLYDHRARAASVSASTTINPYFNATNHTGPM----- 258
Db 320 GEPIGKTSLSLTNKNYENH-----SLSSPOTAINPVFNST--TGPILTLLRRNSSSH 372
QY 259 --BEFGS-YMEGNPRNGSGVKEFEFFPKYGERVS-----VVATYSSLVG--D 302
Db 373 SQKTLGSIQLOQKPR--GSVHK-----PVNRHSRVSSFDKRMESTATAAATVAAYSGSVQ 425
QY 303 CSPNTIDLSL 312
Db 426 LSQNTTPQNL 435

RESULT 7

US-10-007-693-139
; Sequence 139, Application US/10007693
; Patent No. US20020146776A1

; GENERAL INFORMATION:

; APPLICANT: Bhatia, Ajay

; APPLICANT: Probst, Peter

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT

; FILE REFERENCE: 210121.515C2

; CURRENT APPLICATION NUMBER: US/10/007,693

; CURRENT FILING DATE: 2001-12-05

; NUMBER OF SEQ ID NOS: 157

; SEQ ID NO 139

; LENGTH: 660

; TYPE: PRT

; ORGANISM: Chlamydia trachomatis

US-10-007-693-139

Query Match 5.3%; Score 86.5; DB 12; Length 660;
Best Local Similarity 18.9%; Pred. No. 3.7;
Matches 62; Conservative 50; Mismatches 125; Indels 91; Gaps 11;

QY 49 SKTGOQKQKPTLURGMGVAKLERQRIEKKOL-----AAATVGDTRSS 91
Db 291 TKVNLKQEHITGLTDSPLVKKADEQISQAQKDIQIETKPSGSDIPTVGPSSAASAGS 350
QY 92 VASISNATRLPVVD-----PGVVLOGFPSSLSNRNYCGVGGQVMDIP--VISPW 143
Db 351 ALKSSNNGRISLLOOVNEMATAIQGRS-----MIEQFNVNPA 393
QY 144 GFVETSTHESSTINQMFNASS-----NNRCDTCFKKRLDGDQNNVRSNGGFGSKYIMI 189
Db 394 TAKELOAMEAQLTMSD-QLVGADGELPAEIQAIKDALQAALQPSADGLATAMQVAF 452
QY 190 RNSGGSFQYNTIPPPMN-----CYDOYLLQSDHHQSQGL 226
Db 453 AAKVGGSGAGTACTVOMNQYKLYKTAFSTSSSSYAAALSDGYSAY-----KTLNSL 504
QY 227 YDHRARAASVSASTTINPYFNATNHTGPMEEFGSYMEGNPRNGSGVKEFEFFPKY 286
Db 505 YSE--SRSGVQSAISOTANPALSRVSRSG-IESQGRSADASQRAAETIVRDSQTLGDVY 561
QY 287 GERSVSVATYSSLVGDCSPNTIDLSLKL 314
Db 562 SRLQVLDLSMTSTIVSNPQANQBEIMOKL 589

RESULT 8

US-09-801-368-364
; Sequence 364, Application US/09801368
; Patent No. US20020128250A1

; GENERAL INFORMATION:

; APPLICANT: Busby, Robert

; APPLICANT: Cali, Brian

; APPLICANT: Hecht, Peter

; APPLICANT: Holtzman, Doug

; APPLICANT: Madden, Kevin

; APPLICANT: Maxon, Mary

; APPLICANT: Milne, Todd

; APPLICANT: No. US20020128250A1man, Thea

; APPLICANT: Royer, John

; APPLICANT: Salama, Sofie

; APPLICANT: Sherman, Amir

; APPLICANT: Silva, Jeff

; APPLICANT: Summers, Eric

; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fu

; FILE REFERENCE: 109272.147

; CURRENT APPLICATION NUMBER: US/09/801,368

; CURRENT FILING DATE: 2001-03-07

; PRIOR APPLICATION NUMBER: US 09/487,558

; PRIOR FILING DATE: 2000-01-19

; PRIOR APPLICATION NUMBER: US 60/160,587

; PRIOR FILING DATE: 1999-10-20

; NUMBER OF SEQ ID NOS: 440

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 364

; LENGTH: 1250

; TYPE: PRT

; ORGANISM: Saccharomyces cerevisiae

US-09-801-368-364

Query Match 5.3%; Score 86.5; DB 10; Length 1250;
Best Local Similarity 19.9%; Pred. No. 9.5; Mismatches 138; Indels 139; Gaps 17;
Matches 81; Conservative 49;

QY 2 ATSLFFMSTQNSVGNPDLL---RNRLVNVSSSGIRTIETLKSRGKPSGTGQKOKK 58
Db 453 ATSNFLSSPSSSDSLSKDLSVRKRSTINNDSDLSLSPKSGVRRSSLSLKQRTQKK 512
QY 59 ---PTLGMGVAKLERQRIEKK-----QAAATVG---DTSSVASIS 96
Db 513 NDDVEGEGSLLLVEEEIINDKYKPLYAGRVAVLDRIQQLFSOTGLLRPSQOANSN 572
QY 97 NNATRLP-----VPVDPGVVLOGFPSSLG-----SNRYCGVGGQVMDIPVI 140
Db 573 NKPPQSKIAWFPTDKKVPPLIAIPTELAPKDFVENADYSEKLFVASIKRM-----PIT 627
QY 141 S--PWGFVETSTHESSTINQMFNASSNRCDTCFKKRLDGDQNNVRSNGGFGSK 198
Db 628 SLHPFGIL---VSELGLDHP-----DTEIDSLRDN--NFLS 660
QY 199 YTMIPPMNGYDOYLLQSDHHQSQGLFDHRTARAASVSASTTINPYFNATNHTGPM 258
Db 661 ---NEYLQKNPQKPKSPFQPLPLTAESLEYERNFTDTNEYNIFAISELQWV 709
QY 259 EEF-----GSYMEGNPRNGSGV----- 276
Db 710 SEFALHVRNNGNGTLELGCCHVVDVTSHEEGSSVDRARKRSASFMPKOKLVNLLPQSPN 769
QY 277 KEYEFPFGKYGERSVVAT-----TSSLVGD--CSPNTIDLSLKL 313
Db 770 DELSLAPKESATLSVVYTLSDSLTRIKSTWVGESTISPSNI--LSLE 815

RESULT 9

US-09-888-615-64

; Sequence 64, Application US/09888615

; Patent No. US20020064856A1

; GENERAL INFORMATION:

QY	52	-----GQKQKK---PTLRGMGVAKLER-----	QRIEEKKOLA	82
DB	140	ALYVAGEQAQASIADSTLQAGGVRVERGANVTQRTIVDGLHIGTLQLQPLQPEDLPPS	199	
QY	83	AATVGDTSVASISNNAIRLPPVDPGVVLOGFPSSLSNRI-----YCGGVSGOVMI	136	
DB	200	RVVLGDT-----VTAVPASGAPAAV-----SVFGENELTVDGGHITGGAAGVAA	246	
QY	137	D-----PVTSPMCFVETSSITHIEL--SSISNPQ	162	
DB	247	DGAIVHLQPATIRRGDAPAGGAVPGGFCPLLDGHWGVDSUSTVDLAQSIVEAQ	306	
QY	163	MFNASSNRR-COTCFKKKRLDGDQNNVVRNSGGGFSKYTIMPPMNGYQOYLLOSDHQR	221	
DB	307	LGAARFAGRGARVTVSGGSLAPHGNIETGGGA---RRFPPASP-SITLQAGARAQ	361	
QY	222	SGFLX	227	
DB	362	GRALLY	367	
RESULT 11				
US-09-741-669-304				
; Sequence 304, Application US/09741669				
; Patent No. US20020022718A1				
; GENERAL INFORMATION:				
; APPLICANT: Ohlson, R. Allyn				
; APPLICANT: Ohlson, Kari L.				
; APPLICANT: Zyskind, Judith W.				
; TITLE OF INVENTION: Genes identified as required for				
; TITLE OF INVENTION: proliferation of E. coli				
; FILE REFERENCE: ELITRA.009A				
; CURRENT APPLICATION NUMBER: US/09/741,669				
; CURRENT FILING DATE: 2000-12-19				
; PRIOR APPLICATION NUMBER: US 60/173005				
; PRIOR FILING DATE: 1999-12-23				
; NUMBER OF SEQ ID NOS: 481				
; SOFTWARE: FastSeq for Windows Version 4.0				
; SEQ ID NO 304				
; LENGTH: 1325				
; TYPE: PRT				
; ORGANISM: Escherichia coli				
US-09-741-669-304				
Query Match 5.2%; Score 84.5; DB 10; Length 1325;				
Best Local Similarity 22.8%; Pred. No. 16;				
Matches 84; Conservative 41; Mismatches 145; Indels 99; Gaps 21;				
QY	8	MSTDONSNGPN--DLLNRTRLVYNSSGEIR-TETLKSRG-----RKPG	48	
DB	739	MGSEVTIANPCTLDILAST---NSAGDYTLTNALKGDLMRVQLSSDDKMGFTHTATG	794	
QY	49	SK---TGQKQKKPRLRGMGVAKLERQRIEEKKOLAATVGTDTSSVASISNNA-----	99	
DB	795	TEFAGVAQLKDDFTFLERDNTAALTHAMLQSDSENTTSVKVGF-QSTGGIAMNGGTIIFD	853	
QY	100	TRLP-----VPVDPGVVLOGFPSSLSNRIYCGGVSGQVMDPVPSPWG--FVETS	149	
DB	854	TDIPAATLAEGYISVDTLVVGAGDYTWKGRNYQV---NGTGDVLID-VPKPWNDPMANNP	909	
QY	150	STT-----HELSSISNPQMFNASSNRRCDTCFKKRLDGD-----QNNVRSNG	193	
DB	910	LTTLNLEHDDSHV-GVOLVKAQTVIGSGSLTRLRDLQDEVEADKTLHIAQNGTVVAEG	968	
QY	194	G-GFSKYTIMPPMNG-YDOYLLQS-DHHRSQGFLYDHIIRAAVSASS-----TT	243	
DB	969	DYGFRLTT---APGNGLYVNYGLKALNIHGQOKLTLAEHGGAGATADMSKATGGEGDLA	1025	
QY	244	IN-----PYNEATNHHCPEEFSYMEGPNRG-----SGVKEYEF	281	
DB	1026	INTVRQVSLNSGQNDYQGATYYQMGLTRTDADGALGNTRRELNTSNAIIVDLNGSTQVET	1085	

QY 282 FPGKYGERV 290
Db 1086 FTQMGSTV 1094

RESULT 12

US-09-912-020-302

; Sequence 302, Application US/09912020

; Patent No. US20020045592A1

; GENERAL INFORMATION:

; APPLICANT: Zyskind, Judith

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Trawick, John

; APPLICANT: Forsyth, R. Allyn

; APPLICANT: Froelich, Jamie M.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN

; FILE REFERENCE: ELITRA.001DV1

; CURRENT APPLICATION NUMBER: US/09/912,020

; PRIOR FILING DATE: 2001-07-23

; PRIOR APPLICATION NUMBER: 09/492,709

; PRIOR FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: 60/117,405

; PRIOR FILING DATE: 1999-01-27

; NUMBER OF SEQ ID NOS: 485

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 302

; LENGTH: 2383

; TYPE: PRT

; ORGANISM: E. Coli

US-09-912-020-302

Query Match 5.2%; Score 84.5; DB 10; Length 2383;
Best Local Similarity 18.9%; Pred. No. 38;
Matches 77; Conservative 58; Mismatches 147; Indels 125; Gaps 17;

QY 6 FPMSTQNSVGN-----PNDLLNRLVLVNSGSEIRTEFLKSRGKPGSKTQOQ 54
Db 983 YMTATLQKNGNPLDKDEITFSPNDVA--SKFSISNGKGMTD---SNGVAIASLTG-- 1035
QY 55 KQKKPTLRG--MGVAKLERQRIEEK-----KQLAAA 84
Db 1036 -----TLAGTMMIMARLANSVSDAQPMTFVADKDRVAVVVLQTSKAETIGNGVDETTLTA 1090
QY 85 TVGDTSS--VASISNATRLPVPDPGVVLO---GPPSSLSGSRNRYGGVSGQVMIDP 138
Db 1091 TVKQPSNHPVAGITVNTF--MQDVAANETLENNGIAITQANGEAHVTLKGGKAGTHVTA 1149
QY 139 VI-----SPWGFVETSTHSLSSINPOMFNASSNNRCDICFKKRLDQGNVVR 190
Db 1150 TLGNNTSDSOPVTFVADKASQVVLQTSKDEITGNGVDSATLTATVKDQFDNEVNLVP 1209
QY 191 SNGGFSKYTIMPPMNCYDYLQSDHQBOSQFLDHRITARAASVSASSTTINPYNE 250
Db 1210 TFSASSGLTLTPGVSNSTNSGIAQA---TLGVAFECK-----TVTASLANNGASDNK 1260
QY 251 AINHTG-----PMERFGSYMEGNPRNGSGG-----VKEYEFP----- 283
Db 1261 TVHIGDTAAAKIIEAPVDP--SLIAGTFQNSGSGVITATVVDNNGFPVKGVTVNTSN 1318
QY 284 -----GKYGERVSVVATTSSLVGDCSPNTIDLSLK 313
Db 1319 AATAEMTNGQAVTNEQCK--ATVYTNTRSSIESGARPDTVEASLE 1363

RESULT 13

US-09-815-242-5845

; Sequence 5845, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5845
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5845

Query Match 5.1%; Score 83.5; DB 10; Length 1016;
Best Local Similarity 18.9%; Pred. No. 14;
Matches 62; Conservative 58; Mismatches 135; Indels 73; Gaps 14;

QY 12 QNSVGNPNLLNRLVLVNSGSEIRTEFLKSRGKPGSKTQOQ-----QKKPTLRG 63
Db 6 ENNTENTTLKDNVQ-----SKEVKEIEVTKNDTAPOGVEAKSEVTNKKDTIEHPSVKA 60
QY 64 MGVALERQRIEEKKQLAAATVGTSTSVASISNNA-----TRLPPVPDPGV----- 110
Db 61 EDISKK-----EDTPKEV--ADVAEVQPKSSVTHNAETPKVKARSVDSESFIDTRSKN 113
QY 111 VLOGFPSSLSGSRNRYGGVSGQVMIDPVTSPWGFVETSTHSLSSINPOMFNASSNN 170
Db 114 VVESTPTITI--QKKEHFEYGS-----VDIQKKPTDLGVSEVTRFNVGNESNGLIGALQKN 168
QY 171 RCDTC---FKKELDQDNVVSNGG--FSKYTIMPPMNG---YDQYLQSDHQR 221
Db 169 KIDFSKDFNFKVRVANNHQSNTTGADGWGFLFSKGNAAEYLTNGGILGDLGVLNMSGFKI 228
QY 222 SOGFLYDHRITARAASVSASSTTINPYNEATNHTGPMEEFGSYMEGNPRNGSGVKEYEF 281
Db 229 DTGYVYTSNDKTEKQAGQ-----YRGYCAFY-----KNDSSGNSQ--- 265
QY 282 FPGKYGERVSVVATTSSLVGDCSPNTID 309
Db 266 ---MVGENIDKSKTNFLNADNSTNTSD 290

RESULT 14

US-09-925-301-1409

; Sequence 1409, Application US/09925301

; Patent No. US20020052308A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: PA106

; CURRENT APPLICATION NUMBER: US/09/925,301

Db	88	LVSGIPTTAS-----GVMIHQIPQVLPANMAINVOAVNGGMIHAPNAV	133
Qy	135	--MIDPV-----ISPWGFVETSSITHELSSITNPOMFNASSNN	170
Db	134	HPMVPIMAQPAPIHASAASFOPATSPMPISTYTPVPSQFTSFQSSITGSIQ-----SNS	188
Qy	171	RODTQCFKKRLGDQNNVYRSGGGFSKYTMLP-PPMNQY--DQYLLOSDHHQRS-----	222
Db	189	DYSSIFSNMV--RVNTPRS-----VPNSPDGYLHQOHIPQOYQHOTASPSVA	235
Qy	223	-QGFLYDHEIARAAS-----VSASSTII-NPYFNEATNHTGPMEEGCSYMEGNPRGS	273
Db	236	KQOKFAHSLASALSTLQKRTVPVSPATITTESPSSPSDSSTHSASSSALSLPFSNAPSOL	295
Qy	274	GGVKEYE 280	
Db	296	AVAKELE 302	

Search completed: January 26, 2003, 00:07:52
Job time : 25 secs

```

RESULT 15
IS-09-801-368-210
Sequence 210, Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Cali, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: No. US20020128250A1man, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801,368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 210
LENGTH: 382
TYPE: PR1
ORGANISM: Saccharomyces cerevisiae
IS-09-801-368-210

Query Match          5.0%; Score 82.5; DB 10; Length 382;
Best Local Similarity 22.5%; Pred. No. 4;
Matches 69; Conservative 39; Mismatches 99; Indels 101; Gaps 16;

yy      35 IRTEITLKRGRKPG-----SKTQQQKKPYTLRGMGVAKLERORIEEEKQLAA 83
::| :: | || |:: : | : |
bb      36 LRTHT----GKPHCAFPFGCKGSFSDLELKHRRMTHTGQSQRRLKASVQKE----F 87
::| :: | || |:: : | : |
      84 ATVGDTSSVASISNNATPLPVPDPGYVL--QGFPSSLGNS-RIVCGGVGGGV-----134

```

This Page Blank (uspto)

1 protein - protein search, using sw model

run on: January 26, 2003, 00:01:17 ; Search time 41 seconds
(without alignments)
736,249 Million cell updates/sec

File: US-09-701-023-4
Perfect score: 1634
Sequence: 1 MATSLFFMSTQNSVGNPNQ.....TTSSLVGDCSPNTIDLSLKL 314
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Aligned: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 0%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1628	99.6	314	2 T05757	hypothetical prote
2	109	6.7	554	2 S67284	hypothetical prote
3	106	6.5	1027	2 T38759	zinc finger/leucin
4	103	6.3	276	2 T51685	myb-related transc
5	100	6.1	369	2 T01196	transcription fact
6	99.5	6.1	676	2 S61977	transcription fact
7	99.5	6.1	1034	2 T22166	hypothetical prote
8	98	6.0	1480	2 T05566	hypothetical prote
9	97.5	6.0	561	2 T34368	hypothetical prote
10	97.5	6.0	698	2 T32594	hypothetical prote
11	97	5.9	593	2 A36783	unknown protein f2
12	96.5	5.9	669	2 T51246	ARR1 protein [limp
13	96.5	5.9	938	2 S20480	trag protein - Esc
14	95.5	5.8	342	2 AG1729	protein gp19 (Bact
15	95.5	5.8	894	2 S51245	probable finger pr
16	95.5	5.8	1022	2 T24663	hypothetical prote
17	95.5	5.8	1165	2 S62982	vacuolar protein v
18	95.5	5.8	2232	2 T34434	hypothetical prote
19	95	5.8	302	2 H96792	unknown protein f1
20	95	5.8	625	2 S48941	regulatory protein
21	94	5.8	461	2 T00621	hypothetical prote
22	94	5.8	1777	2 T34369	hypothetical prote
23	93.5	5.7	444	2 B90053	hypothetical prote
24	93.5	5.7	1091	2 S57112	USM1 protein - yea
25	93	5.7	655	1 A55726	RNA-binding protei
26	93	5.7	656	1 A49358	RNA-binding protei
27	92.5	5.7	371	2 D90192	alcohol dehydrogen
28	92.5	5.7	601	2 T11677	probable transcrip
29	92	5.6	290	2 T05009	hypothetical prote

30 92 5.6 419 2 T18420 hypothetical prote
31 91.5 5.6 434 2 T01013 hypothetical prote
32 91 5.6 788 1 J0VLRH DNA-directed DNA p
33 91 5.6 1424 2 T03851 thyroid hormone re
34 90.5 5.5 288 2 S33714 ribosomal protein
35 90.5 5.5 632 2 T02627 hypothetical prote
36 90.5 5.5 959 2 B44402 nuclear pore compl
37 90.5 5.5 967 2 S66852 hypothetical prote
38 90.5 5.5 1163 2 A36685 205K microtubule-a
39 90 5.5 828 2 C88402 protein H05C05.1 f
40 90 5.5 872 2 T37789 Scd1 protein - fis
41 90 5.5 914 2 S46593 finger protein AZF
42 90 5.5 1030 2 F96763 hypothetical prote
43 90 5.5 1203 2 T01287 hypothetical prote
44 90 5.5 1970 2 T03284 myoblast city prot
45 89.5 5.5 454 2 T47253 core protein II [i

ALIGNMENTS

RESULT 1

T05757
hypothetical protein M4122.140 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress);
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 04-Mar-2000
C:Accession: T05757
R:Bevan, M.; Reichert, B.J.; Barel, E.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.X.; Sch
submitted to the Protein Sequence Database, June 1998
A:Reference number: Z15450
A:Accession: T05757
A:Molecule type: DNA
A:Residues: 1-314 <BEV>
A:Cross-references: EMBL:AL030978
A:Experimental source: cultivar Columbia; BAC clone M4122
C:Genetics:
A:Map position: 4
A:Introns: 177/3; 255/3
A:Note: M4122.140
C:Superfamily: Arabidopsis thaliana hypothetical protein M4122.140

Query Match 99.6%; Score 1628; DB 2; Length 314;
Best Local Similarity 99.7%; Pred. No. 4.3e-127;
Matches 313; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MATSLFFMSTQNSVGNPNQDLNRLVNVSSGEIRTEFLKSRGKPGSKTGQOKKPT 60
Db 1 MATSLFFMSTQNSVGNPNQDLNRLVNVSSGEIRTEFLKSRGKPGSKTGQOKKPT 60
QY 61 LRGMGVAKLERQRIEERKQLAAATVGDFTSSVASISNNATRLPVPVDPGVVLOGFPSSLG 120
Db 61 LRGMGVAKLERQRIEERKQLAAATVGDFTSSVASISNNATRLPVPVDPGVVLOGFPSSLG 120
QY 121 SNRIYCGGVGSGQVMDIPVISPWGFVETSTTHELSSISNPQMFNASSNNRCDTCFKKR 180
Db 121 SNRIYCGGVGSGQVMDIPVISPWGFVETSTTHELSSISNPQMFNASSNNRCDTCFKKR 180
QY 181 LDGDQNNVRSNGGFGSKYTMIPPMNGYDQVLLQSDHRSQGLYDHRIRAAASVSAS 240
Db 181 LDGDQNNVRSNGGFGSKYTMIPPMNGYDQVLLQSDHRSQGLYDHRIRAAASVSAS 240
QY 241 STTINPYFNEAHTGPMEEFGSYMEGNPRNGSGGKEYEFPFGKYGERVSVVATSSIV 300
Db 241 STTINPYFNEAHTGPMEEFGSYMEGNPRNGSGGKEYEFPFGKYGERVSVVATSSIV 300
QY 301 GDCSPNTIDLSLKL 314
Db 301 GDCSPNTIDLSLKL 314

RESULT 2
S67284
hypothetical protein YOR372c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein O6682

C:Species: Saccharomyces cerevisiae

C>Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 19-Apr-2002

C:Accession: S67284

R:Delius, H.; Hebling, U.; Hofmann, B.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67261

A:Accession: S67284

A:Molecule type: DNA

A:Residues: 1-554 <DE>

A:Cross-references: EMBL:Z75280; NID:g1420803; PID:e252195; PID:g1420804; GSPDB:GN00015

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:NDJ1; MIPS:YOR372c

A:Cross-references: SGD:S0005899

A:Map position: 15R

Query Match 6.7%; Score 109; DB 2; Length 554;

Best Local Similarity 22.7%; Pred. No. 0.48;

Matches 63; Conservative 37; Mismatches 79; Indels 98; Gaps 13;

QY 7 FNSTD-QNSVGNPDLNRLTRLVNSSGEIRTELKSRGKPCGKTGQKQKPTLRMG 65

Db 71 FANTDAHNSSESLVENSIL-----PHTQQIQOQOQ----- 103

QY 66 VAKLERORIEEEKKOLA-----AATVGTSSVASISNNATRLPVPVDPGVWLGFPSSIG 120

Db 104 -----QQQQQQQQQALGSLVPAVTRTSETLDDIN-----VQPSVYLQ-FGNSLP 150

QY 121 SNRIYCGVGGSGQVMID-----PVISPMGFVETSTTHELSSINP-QMFNAS 167

Db 151 SEELVASPEQFEFLDSPNFNFHTKPTAKTPLRFV-TDSNGAQQTENPGQQQVNF 209

QY 168 SNRCDTCFKKRLGDDGNVVRSG-----GGFSKYTIMPPPMNGYDOYLQSDHH 219

Db 210 SN-----VDLNNLLKNGKTPSSCTGAFSR-----TPLSKIDNML----- 245

QY 220 QRSQGLFYDHRTARAASVSASTTINPVENEATHTG 256

Db 245 -----MFNOPLTSPKRFSSLSLTPYGRKILNDVG 276

RESULT 3

I38759

zinc finger/leucine zipper protein - human

C:Species: Homo sapiens (man)

C>Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 05-Nov-1999

C:Accession: I38759

R:Chaplin, T.; Ayton, P.; Bernard, O.A.; Saha, V.; Della Valle, V.; Hillion, J.; Gregori

Blood 85, 1435-1441, 1995

A:Title: A novel class of zinc finger/leucine zipper genes identified from the molecular

A:Reference number: I38759; MUID:95195207; PMID:788665

A:Status: preliminary

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1027 <RES>

A:Cross-references: EMBL:U13948; NID:g538276; PIDN:AAA79972.1; PID:g538277

C:Genetics:

A:Gene: AF10

Query Match

Best Local Similarity 6.5%; Score 106; DB 2; Length 1027;

Matches 62; Conservative 40; Mismatches 106; Indels 60; Gaps 11;

QY 30 NSSGEIRTELKSRGKPCGKTG-----QQQKQKPTLRGMGVAKLERQRIEKKQ 80

Db 435 NSPDGLNSSLPTAGYKKAQTSIGIEETVTEKKRGNQKSHGPRGPKGNKN--QENVSH 492

QY 81 LAAATVGTSSVASISNNATRLPVPVDPGVVLOGFPSSSLGNSNRIYCGVGGSGQVMIDPVI 140

Db 493 LSVSSASPTSSVAGAAGSITSSSLQKSPTLRLNGSLQSL-----SVGS----- 535

Db 536 SPVGSEISMQRHEDGACPTTTFSELLNAIHDRGDSSTLTQELKFIGIYNSNDVAVSFP 595

QY 187 NVVRNNGGFSKYTIMPPPMNGYDOYLQSDHHQSRQSGFLYDHRIRARAASVSASTTNP 246

Db 596 NVYVSGS-----SSTPVSSSHLPQ---QSSGHLQOVGALSFSVSSAAPAVATT----- 641

QY 247 YFNEATNHTG-PMEEFGSVMEGNPRNGS 273

Db 642 ---QANTLSSGSLSQAPSHMYGNRSNS 666

RESULT 4

T51685

myb-related transcription factor MYB86 [imported] - Arabidopsis thaliana (fragment)

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 17-Nov-2000

C:Accession: T51685

R:Kranz, H.D.; Denekamp, M.; Greco, R.; Jin, H.; Kranz, H.D.; Denekamp, M.; Greco, I

Plant J. 16, 263-276, 1998

A:Title: Towards functional characterisation of the members of the R2R3-MYB gene fr

A:Reference number: Z14349; MUID:9839469; PMID:9839469

A:Accession: T51685

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-276 <KRA>

A:Cross-references: EMBL:AF062913; PIDN:AAC83635.1

A:Experimental source: cultivar Columbia

C:Genetics:

A:Gene: MYB86

C:Superfamily: unassigned myb DNA-binding repeat proteins; myb DNA-binding repeat

C:Keywords: transcription factor

Query Match 6.3%; Score 103; DB 2; Length 276;

Best Local Similarity 23.3%; Pred. No. 0.57;

Matches 60; Conservative 41; Mismatches 107; Indels 50; Gaps 12;

QY 25 TRLVNVSSGEIR---TETLKSRGRKPGSKTQKQKPTLRGMGVAKLERQRIEKKQ 81

Db 18 TRLPGRTDNEIKFNWNSCLKKLRKGI---DPTTHKPLITN-----ELQSLNVIDOKL 68

QY 82 AAATVGTSSVASISNNATRLPVPVDPGVVL---QG---FP---SSLGSNRIYCGVSG 132

Db 69 TSSEV--VKSTGSIINLH-----DOSMVYSSQOOGPMWFPANTTTTQNSAFCSST 119

QY 133 QVMIDPVTSPMGFVETSTTHELSSINPQMFENASSNNRCDFCKKKRLDGDNNVRSN 192

Db 120 TVSDQIVSLSSMSTSSPTPMTSNTSPAPNNEQLNYCNTV-----PSQNSNIYSA 172

QY 193 GGFSKYTIMPPPMNGYDOYLQSDHHQSRQSGFLYDHRIRARAASVSASTT----- 244

Db 173 FFG-NQYTEASQTNMNNNNPLVDQHHHQMDSWASE--ILHYTERSSQSETVIEAEVKPD 229

QY 245 --NPYFNEATNHTGPMEE 260

Db 230 IANYWRSASSSSSPNQE 247

RESULT 5

T01196

transcription factor MYB4 - Arabidopsis thaliana

N:Alternate names: protein F21E10.4

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 19-May-2000

C:Accession: T01196

R:Davidson, S.; Rohlfing, T.; David, M.; O'Brian, D.

submitted to the EMBL Data Library, April 1998

A:Description: The sequence of A. thaliana F21E10.

A:Reference number: Z14258

A:Accession: T01196

A:Status: translated from GB/EMBL/DBJ

[illegible]

Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Mafti, R.; Marziani, M.; Rooney, T.; Rowley, D.; Sakano, H.
Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
Reference number: A86141; MUID:21016719; PMID:11130712
Accession: A96783
Status: preliminary
Molecule type: DNA
Residues: 1-593 <STO>
Cross-references: GB:AE005173; NID:g10092269; PIDN:AGI12682.1; GSPDB:GN00141
Genetics:
Gene: F22H5.2
Map position: 1
Query Match 5.9%; Score 97; DB 2; Length 593;
Best Local Similarity 21.4%; Pred. No. 5.2; Indels 130; Gaps 18;
Matches 78; Conservative 44; Mismatches 112
Y 23 RNT--RLVYSSGEITETLKSRRKPGSKTGQOK-----OKPTLRGMGVAKL 69
b 162 RNTAARLVVQENKKVLPE---GSGKSAANQOQKTEETKALQERKKVLHDDGVQKL 218
Y 70 EROR-----IEEKKOLAA-----ATVGDTSVASISNNATRLPVPVPGVWL 112
b 219 EADGQOKSKETEKALQETKRSLOAVGREDASTRKTRHMAAASETTTRPRDL----- 272
Y 113 QGPPSSLSNRIYCGVGSGQVMDPVISPWGFEVSSTT-----HE-----LSSIS 159
b 273 -----PEKTTQNTQRTIPTDDHOKTKGALTSNLG 303
Y 160 NPOMFNASSNNRCDTCFKKRLDQDNVVRSGGFSKYTMIPPPMNGYDQ---YLLQ 215
b 304 NRVNTNREGSSSRKTEIDRD-----GSKLTV--GKSGDDKSVSVTTLT 349
Y 216 SDH-----HRSQGLYDHRIRAAASVSASSTTI---NPYFNEATNHTGPMEEF 261
b 350 GENKGATMGISEKDKDGEVHIRGVRNPDSSNTATETENPKDDAEAEA---SF 405
Y 262 GSYMEGNPR--NGS---GGVXEYEFPPG-----KYGRVSVVATTSSSLVGDSPNTIDL 310
b 406 TAYINGNTQGINNSIVESSVSEND--PGVHMSFKPILKKEVIYPENVEKKPPTVTV 463
Y 311 SLXL 314
b 464 TKKL 467
RESULT 12
51246
rRI protein [imported] - Arabidopsis thaliana
Species: Arabidopsis thaliana (mouse-ear cress)
Date: 28-Jul-2000 #sequence_revision 28-Jul-2000 #text_change 08-Dec-2000
Accession: T51246
Sakai, H.; Aoyama, T.; Bono, H.; Oka, A.
Plant Cell Physiol. 39, 1232-1239, 1998
Title: Two-component response regulators from Arabidopsis thaliana contain a putative
Reference number: Z25340
Accession: T51246
Status: preliminary; translated from GB/EMBL/DBJ
Molecule type: DNA
Residues: 1-669 <SAK>
Cross-references: EMBL:AB016471; PIDN:CAB19446.1
Experimental source: cultivar Columbia
Genetics:
Gene: ARRL
Introns: 63/1; 114/1; 261/1; 286/3
Superfamily: response regulator homology
Query Match 5.9%; Score 96.5; DB 2; Length 669;
Best Local Similarity 23.7%; Pred. No. 6.7; Indels 85; Gaps 21;
Matches 79; Conservative 51; Mismatches 118

QY 15 VGNPDLNRLNRLVNVSSGEIET-----ETLKSRRKPGSKTGQOKKKKPTLRGM-GVAKL 69
Db 350 LGRP-AMVSKGLPVSSIVDERSIFSDFNTKTRFEGSLGHGQPOQPOQMNLLHGVP TG 408
QY 70 ERORIE-----EEKKOLAAATVGDTSVASISNNATRLP-----VPDPGVVLOGFPSS 118
Db 409 LQOQLPMGNRMSIQOQIAAVRAGN-----SVQNGMLMPLAGQQSLPRGPPMLTSSQSS 463
QY 119 LG-----SNRI--YCGVGSGQVMDPVISPWGFEVSTSTHELSSIS-----NPQM-----F 164
Db 464 IROPMLSNRISERSGSGRNIPESRVLPTSYTNL--TTOHSSSSMPYNNFQELPVNSF 522
QY 165 NASSNNRCDTCFKKRLDQDNVVRSGGFSKYTMIPPPMNGYDQYLL-QSDHHORSQ 223
Db 523 PLASAPGISVPRKAT---SYQEEVNSSEAGF-----ITP---SYDMFTTRQNDWDLRNI 571
QY 224 GFLYD-HRIARAASVSAS-----STTINPYFNEATNH-----TGPMEEFSS 263
Db 572 GIADFQDSESAFSAEYSSSSMSRHNTTV-----AATEHRNHQPPSGMVQHQQV 626
QY 264 YMEGNPRNGSGVKEVEFFPGKYGRVSVVATT 296
Db 627 YADGN--GGSVRVK-----SERVATDTAT 648
RESULT 13
520480
traG protein - Escherichia coli plasmid F
C:Species: Escherichia coli
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S20480
R:Birth, N.; Skurray, R.
Mol. Gen. Genet. 232, 145-153, 1992
A:Title: Characterization of the F plasmid bifunctional conjugation gene, traG.
A:Reference number: S20479; MUID:92204127; PMID:1348105
A:Accession: S20480
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-938 <FIR>
A:Cross-references: EMBL:M59763; NID:g148630; PIDN:AAA98081.1; PID:g148633
C:Genetics:
A:Genome: plasmid F
A:Start codon: GTG
Query Match 5.9%; Score 96.5; DB 2; Length 938;
Best Local Similarity 20.2%; Pred. No. 11;
Matches 66; Conservative 46; Mismatches 153; Indels 61; Gaps 10;
QY 2 ATSLFFMSTDNQSVGNENDLLNRLNRLVNVSSGEIETETLKSRRKRP-GSKTGQOKKKPT 60
Db 479 STTSFGQMYGTGGATATQTRDGNVWMDASG-----AMSRLPVGINATRO----- 524
QY 61 LRGMGVAKLRQRIEEREKKOLAAATVGDTSVASISNNATRLPVPDPGVVLOGFPSSLG 120
Db 525 -----IAAAQOEMAREASNAESALHGFSSSIASAWN-----TLQFGSNRG 566
QY 121 SNRIYCGVGSGQVMDPVISP--WGFVETSTHELSSISNPQMFNASSNNRCDTCFKK 178
Db 567 SDSLVTGGADSTMSAQDSMAASMRSAVESYAKAHNTSNEQATRELASRTNASLGLYGD 626
QY 179 KRLDQDNVVRSGGFSKYTMIPPPMNGYDQYLLQSDHHQRSQGLYDHRIRARAASVS 238
Db 627 AYAKGHGISVLNGGGVGLQAGAKASIDGSD-----LDSHEASSGSRASHADRHIDAR 681
QY 239 ASSTTINPYFNEATNH--TGPMEEFSGYMEGNP-----RNGSGGVKEVEFFPGK 285
Db 682 ATQD-----FREASDYFTSRKVSSEGSHTDNNADSRVDQLSAALNSAKQSYDQTTNMT 736
QY 286 YGERVSVVATTSSLVGDCSPNTIDL 311
Db 737 SHEYAEMASRTESMGQMSQSE---DLS 759

RESULT 14

AG1729
C:Species: *Listeria innocua*
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AG1729
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krest, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mak, C.; Schluster, T.; Simoes, N.; Trierz, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.
A:Reference number: AB1077; MUID:2153/279; PMID:11679669
A:Accession: AG1729
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-342 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC97607.1; PID:g16414903; GSPDB:GN00178
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: lin2380

Query Match 5.88; Score 95.5; DB 2; Length 342;
Best Local Similarity 22.3%; Pred. No. 3.2;
Matches 57; Conservative 35; Mismatches 79; Indels 85; Gaps 11;

```
QY 23 RNTRLVNSGEIRTEGLKSRGKPGSKTGQOKKPTLRGMGVAKLERIEEKKOLA 82
Db 44 RITNLVIASGGSSNEVDLR-----VSKLQNKIFELAKORLD 81

QY 83 AAVYGDTSVA-SISNNATRLPVDPGVVLQGFPPSSILGSNRIYCGGVGQVMIDPVIS 141
Db 82 S-----DLDSLADSLKNNWTRI-----TSIELTNEQVLYMLNRYGLDAGSIEVYVDSVSG 132

QY 142 PWGFEVTSST-----THELSISNPQMFNASSNNRCDTCFKKRLDGDQNNVRSNG 193
Db 133 D-----DTAGTCEKNKPKTKATMNPFRFN---SNTLRWINPGRYDED----- 176

QY 194 GFSKYTMIPPMNGYQYLQSDHH-----ORSQGFLYDRIARAA 235
Db 177 -----VIIPPLUGVTLYLSSNYETVPDPAAGTTQIRSIYSVTSQTVY---IAGIE 226

QY 236 SVSASSTTINPYNEA 251
Db 227 QTNTAGTTKN-YPIKA 241
```

RESULT 15

S51245
N:Alternate names: probable finger protein YDR096w - yeast (*Saccharomyces cerevisiae*)
C:Species: *Saccharomyces cerevisiae*
C:Date: 23-Feb-1995 #sequence_revision 12-May-1995 #text_change 06-Feb-1998
C:Accession: S51245
R:Murphy, L.; Harris, D.
submitted to the EMBL Data Library, January 1995
A:Reference number: S51243
A:Accession: S51245
A:Molecule type: DNA
A:Residues: 1-894 <MUR>
A:Cross-references: EMBL:247746; NID:g6333627; PID:g6333630; MIPS:YDR096w
C:Genetics:
A:Gene: SGD:GIS1
A:Cross-references: SGD:S0002503; MIPS:YDR096w
A:Map position: 4R
C:Keywords: zinc finger

Query Match 5.88; Score 95.5; DB 2; Length 894;
Best Local Similarity 20.7%; Pred. No. 12;
Matches 69; Conservative 42; Mismatches 135; Indels 87; Gaps 15;

```
QY 12 QNSVGNPDLLRNLRLVNVSSGEIRT-----ETLKSRRKP--GSKTGQOKQ 56
Db 378 ONLEDNKNPLFSNINMNRPOSSLSRSTTPGVNQVFLNNQTTISRISPLSRMMDLSNI 437

QY 57 KKPTLRGMGVAKLERIEEKKOLAATVCDTSSVASISNNATRL-----PVPVDPGVWL 112
Db 438 VETILDPPG-SKPKRKVLITQLQMNIPSSNSNFGTSPSLINTNSLLSNITATINPSTTT 496

QY 113 QGPPSSLGSRNRIYCGGVGQVMIDPVISPWGFVETSSSTTHELSSISNPQMFNASSNNRC 172
Db 497 NG---SQNHNNVNANGINTS-----AAASINNNISSTNN-SANNSSSSNNV 538

QY 173 DTCEKKRLDGDQNNVRSNGGFSKYTMIPPPMNGYDQYLQSDHHQSQGFLYDRIAR 232
Db 539 STV-----PSSMMHSS-----TLNGTSG--LGGDND-----DNMLA 567

QY 233 RAASVSASSTTINPY-----FNEATNHTGPMEEFGSYMEGNPRNGSGGVKEYEFFFPGKYG 267
Db 568 LSLATLANSATASPRLTLPPLSSPMNPNGHTSYNGNMNNNSNGSGNSNSY-----SNG 622

QY 288 ERVSVVATTS-----SLVG---DCSPNTIDL 312
Db 623 VTAAATTTTAPHNLSIVSPNPTYSNPPLSLYL 655
```

Search completed: January 26, 2003, 00:06:44
Job time: 51 secs

	(without alignments)	updates/sec
591.980 Million cell		

```

effect score: 1634
sequence: 1 MATSLFFMSTDONSVGNPD.....TTSSLVGDSPNTIDISLKL 314

```

Gapop 10.0 , Gapext 0.5

total number of hits satisfying chosen parameters: 112892

Maximum DB seq length:	2000000000
Maximum DB seq length:	2000000000

Maximum Match 100%

atabase :
SwissProt 40:*

SUMMARIES

result No.	Score	Query Match	Length	DB	ID	Description	
1	116.5	7.1	1068	1	AF10_MOUSE	O54826	mus musculus
2	106	6.5	1027	1	AF10_HUMAN	P55197	homo sapien
3	99.5	6.1	676	1	RLM1_YEAST	Q12224	saccharomyc
4	96.5	5.9	938	1	TRG1_ECOLI	P33790	escherichia
5	95.5	5.8	389	1	SER1_BOMBO	P07856	bombyx mori
6	95.5	5.8	1165	1	YNF4_YEAST	P33950	saccharomyc
7	95	5.8	625	1	R101_YEAST	P33400	saccharomyc
8	93.5	5.7	1091	1	JSN1_YEAST	P47135	saccharomyc
9	93.5	5.7	1516	1	NCO2_XENLA	Q9W705	xenopus lae
10	93	5.7	655	1	EMS_MOUSE	Q61545	mus musculu
11	93	5.7	656	1	BWS_HUMAN	Q01844	homo sapien
12	91	5.6	788	1	DPOL_HPBBE	P13846	heron hepat
13	90.5	5.5	959	1	N100_YEAST	Q02629	saccharomyc
14	90.5	5.5	1185	1	MAPX_DROME	P32226	drosophila
15	90.5	5.5	1464	1	NCO2_HUMAN	P55596	homo sapien
16	90	5.5	279	1	T0NB_HABSD	Q51810	haemophilus
17	90	5.5	872	1	SCD1_SCHPO	P40995	schizosacch
18	90	5.5	914	1	AZF1_YEAST	P41696	saccharomyc
19	90	5.5	1338	1	PUR4_HUMAN	Q15067	homo sapien
20	89.5	5.5	454	1	UCR2_NEUCR	O60044	neurospora
21	89.5	5.5	647	1	VE1_HPV40	P36727	human papil
22	89	5.4	1230	1	ST20_CANAL	Q92212	candida alb
23	88.5	5.4	954	1	FLEY_CAOCR	P15345	caulobacter
24	88.5	5.4	1170	1	YKD8_YEAST	P32862	saccharomyc
25	88	5.4	1640	1	SSRP_ARATH	Q05153	arabidopsis
26	88	5.4	1113	1	N116_YEAST	Q02630	saccharomyc
27	87.5	5.4	832	1	YFC4_YEAST	P43572	saccharomyc
28	87	5.3	1462	1	NCO2_MOUSE	Q61026	mus musculu
29	87	5.3	3587	1	SRF2_BACSU	Q04747	bacillus su
30	86.5	5.3	561	1	YD89_SCHPO	Q10414	schizosacch
31	86.5	5.3	566	1	ATF1_SCHPO	P32890	schizosacch
32	86.5	5.3	1250	1	SSD1_YEAST	P24276	saccharomyc
33	86.5	5.3	1465	1	NCO2_RAT	Q9WU19	rattus norv

QY 64 MGAKLPRRIEKKOLAATVGTSSVASISNNATRLPVPDGVVLQGFSSLSGSR 123
 Db 478 PGRPKGNKN--QENVSHLVSSASPTSSVASAGSVTSSSLOKSPPTLLRNGSLQSL----- 531
 QY 124 IYCGVGSGQVMDPVTSPGKFTVSTTHLSSISNPQMFNASSNRCDTCFKKKRLDG 183
 Db 532 ----SVGS-----SPVGSEISMQRHD-----GACPTTFSELLNA 563
 QY 184 DONNVRSNGGFSKYTMIPPMNGYDQYLQSDHH--QRSOGFLYDHRRTARASVSASS.241
 Db 564 IHNGIYNSNDVAVS-----FPNVVSGSGSTPVSSSHPTQOSSGHL--QQVGALSPSAASS 617
 QY 242 TTINPYNEATNHTG-----PMEEFSGSYMEGPN 269
 Db 618 VTPAAATTOANTVSGSSLSQAPAHMYSRLNQP 651

RESULT 2
 ID AF10_HUMAN STANDARD; PRT; 1027 AA.
 AC P55197;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE AF-10 protein.
 GN MLLT10 OR AF10.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95195207; PubMed=7888665;
 RA Chaplin T., Ayton P., Bernard O.A., Saha V., della Valle V.,
 RA Hillion J., Gregorini A., Lillington D., Berger R., Young B.D.;
 RT "A novel class of zinc finger/leucine zipper genes identified from
 RT the molecular cloning of the t(10;11) translocation in acute
 RT leukemia";
 RL Blood 85:1435-1441(1995).
 CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
 CC -1- TISSUE SPECIFICITY: ABUNDANTLY IN TESTIS.
 CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY A CHROMOSOMAL
 CC TRANSLOCATION T(10;11)(p12;q23) THAT INVOLVES MLLT10 AND MLL/HRX.
 CC THE RESULT IS A ROGUE ACTIVATOR PROTEIN.
 CC -1- SIMILARITY: HIGH, TO AF17.
 CC -1- DATABASE: NAME-Atlas Genet. Cytoenet. Oncol. Haematol.;
 CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/AF10.html".
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U13948; AAA79972.1; -
 CC Genew; HGNC:16063; MLLT10.
 CC MIM; 602409; -
 CC InterPro; IPR001965; znf_PHD.
 CC Pfam; PF00628; PHD; 1.
 CC SMART; SM00249; PHD; 2.
 CC PROSITE; PS01359; ZF_PHD.1; 1.
 CC PROSITE; PS00016; ZF_PHD.2; 2.
 CC Nuclear protein; Zinc-finger; Repeat; Chromosomal translocation;
 KW Proto-oncogene.
 KW ZN_FING 22 74 PHD-TYPE 1.
 FT ZN_FING 22 85 C4-TYPE.
 FT ZN_FING 134 197 PHD-TYPE 2.
 FT ZN_FING 229 240 PHD-TYPE 2.
 FT DOMAIN 229 240 GLU/LYS-RICH.
 FT DOMAIN 766 794 LEUCINE-ZIPPER.

FT DOMAIN 856 861 POLY-SER.
 FT SITE 266 266 MLL FUSION POINT (IN ACUTE MYELOID
 FT LEUKEMIA PATIENT B).
 FT SITE 643 643 MLL FUSION POINT (IN ACUTE MYELOID
 FT LEUKEMIA PATIENT C).
 FT SITE 680 680 MLL FUSION POINT (IN ACUTE MYELOID
 FT LEUKEMIA PATIENT A).
 SQ SEQUENCE 1027 AA; 109026 MW; 7C7C0059DD46589 CRC64;
 Query Match 6.5%; Score 106; DB 1; Length 1027;
 Best Local Similarity 23.1%; Pred. No. 0.94;
 Matches 62; Conservative 40; Mismatches 106; Indels 60; Gaps 11;
 QY 30 NSSGEIRTEITLKSRGRKPGSKTG-----OQKQKPTLRGMGVAKLRFRIEKKQ 80
 Db 435 NSPGDLGNSLPTAGYKRAQTSGLIEETVKKRKKGNKQSKHGPRKGNK--QENVSH 492
 QY 81 LAAATVGTSSVASISNNATRLPVPDGVVLQGFSSLSGSRNIYCGVGSGQVMDPVI 140
 Db 493 LSVSSASPTSSVASAAGSISSLSQKSPPTLLRNGSLQSL-----SVGS----- 535
 QY 141 SPWGFVETSTTHE--LSSISNPQMFNASSNRCDTCFKKK-----RLDGDQ-- 186
 Db 536 SPVGSEISMQRHDGACPTTFSELLNATHNRDGSSTLTQKELKFTGIYNSNDVAVSFP 595
 QY 187 NVVRSNGGFSKYTMIPPMNGYDQYLQSDHHQRSGFLYDHRIRARASVSASSTINP 246
 Db 596 NVVSGSG-----SSTPVSSSHLPQ---QSSHLQOVLGALSPSAVSAAPAVTT----- 641
 QY 247 YFNEATNHTG--PMEEFSGSYMEGPNRNGS 273
 Db 642 ---QANTLSSLSQAPAHMYSRLNQP 656

RESULT 3
 RLML_YEAST
 ID RLML_YEAST STANDARD; PRT; 676 AA.
 AC Q12224;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcription factor RLML.
 GN RLML OR YPL089C OR LFC19C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetaceae;
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=96009602; PubMed=7565726;
 RA Watanabe Y., Irie K., Matsumoto K.;
 RT "Yeast RLML encodes a serum response factor-like protein that may
 RT function downstream of the Mpk1 (Slf2) mitogen-activated protein
 RT kinase pathway";
 RL Mol. Cell. Biol. 15:5740-5749(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=97313271; PubMed=9169875;
 RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansorge W.,
 RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
 RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
 RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
 RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
 RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
 RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,
 RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
 RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
 RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
 RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
 RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
 RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,

TRG1_ECOLI STANDARD; PRT; 938 AA.

P33790;
01-FEB-1994 (Rel. 28, Created)
01-FEB-1994 (Rel. 28, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Trag protein.
TRAG.
Escherichia coli.
Plasmid F.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
NCBI_TaxID=562;
[1]
SEQUENCE FROM N.A.
STRAIN-K12;
MEDLINE=92204127; PubMed=1348105;
Firth N., Skurray R.A.;
"Characterization of the F plasmid bifunctional conjugation gene,
trag.";
Mol. Gen. Genet. 232:145-153(1992).
[2]
SEQUENCE FROM N.A.
MEDLINE=94359430; PubMed=7915817;
Prost L.S., Ippen-Ihler K., Skurray R.A.;
"Analysis of the sequence and gene products of the transfer region of
the F sex factor.";
Microbiol. Rev. 58:162-210(1994).
[3]
SEQUENCE FROM N.A.
STRAIN-K12 / CR63;
Shimizu H., Saitoh Y., Suda Y., Sampei G., Mizobuchi K.;
"Complete nucleotide sequence of the F plasmid: its implications for
organization and diversification of plasmid genomes.";
Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: PLAYS A CRUCIAL ROLE IN DONOR-RECIPIENT CELL
INTERACTIONS. REQUIRED FOR TWO STAGES OF THE CONJUGATION PROCESS:
PILUS BIOSYNTHESIS AND MATING AGGREGATE STABILIZATION. MAY
INTERACT WITH TRAN.
-!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
-!- PTM: TRAG* PROBABLY ARISES FROM THE POST-TRANSLATIONAL CLEAVAGE OF
TRAG.
-!- CAUTION: TRAG IS NOT RESPONSIBLE FOR THE N-TERMINAL ACETYLATION
OF F PILIN AS STATED BY SOME AUTHORS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M59763; AAA98081.1; -.
EMBL; U01159; AAC44184.1; -.
EMBL; AP001918; BAA97969.1; -.
FIR; S20480; S20480.
DR Ecogene; EG40114; trag.
KW Plasmid; Conjugation; Transmembrane; Inner membrane;
Complete proteome.
FT CHAIN 1 938
FT CHAIN ? 938
FT CHAIN ? 938
FT DOMAIN 1 53
FT TRANSMEM 54 73
FT DOMAIN 74 329
FT TRANSMEM 330 350
FT DOMAIN 351 412
FT TRANSMEM 413 433
FT DOMAIN 434 938
SQ SEQUENCE 938 AA; 102471 MW; 9E09C8402ACDFDDF CRC64;

RESULT 4

Matches	56;	Conservative	46;	Mismatches	153;	Indels	61;	Gaps	10;
QY	2	ATSLFFMSTDQNSGVGNPDILLNTRLVWSSNGSIRTEITLKSGRKP--GSKTGQOQKPKT	60						
		: : :	: : :	:	:	:	:	:	:
Db	479	STTSFGOMVYQTGSGATATOTDRGNVMDASG-----AMSLRPLVGINATRQ-----	524						
QY	61	LRGMGVAKLQRRIEKKQLAAATYGDTSVVASISNNATRLPVPDPGVVLQGFPSLIG	120						
		: : :	: : :	:	:	:	:	:	:
Db	525	-----TAAAOQEMAREASNRASALHGFSSSIASAWN-----TLSQSGNRG	566						
QY	121	SNRIYCGGVSGQVMIDPWLSP--WGFVETSTTHLSSISNPDMFNASSNRCDTCFKK	178						
		: : :	: : :	:	:	:	:	:	:
Db	567	SSDSVTGGADSTMSAODSMMSMRGSAVESAKAHNISNEQATRELASRSTNASLGLYGD	626						
QY	179	KRLDGQNNVVRNNGGGFSKYTIMIPPMNGYDQYLLQSDPHORSOGFLYDHRIRAAASY	238						
		: : :	: : :	:	:	:	:	:	:
Db	627	YAKGHLGTVLNGGVGLQAGAKASIDGSD-----LDSHSAAGSRASHDARHDIDAR	681						
QY	239	ASSTINPYFNEATNH--TGPMEEFQSYMEGNP-----RNGSGGVKEYEYFFPGK	285						
		: : :	: : :	:	:	:	:	:	:
Db	682	ATOD----FKEASDYFTSRKYSSESGSHDNDADSDVOLSAAALSAKQSYDQYTYTNMTR	736						
QY	286	YGERVSVWATTSVLGDCSPNTIDLS	311						
		: : :	: : :	:	:	:	:	:	:
Db	737	SHEYAEWASRTESMSGOMSE---DLS	759						

RESULT 5

```

SERI_BOMMO
ID  SERI_BOMMO  STANDARD;  PRT;  389  AA.
AC  P07856;
DT  01-AUG-1988  (Rel. 08, Created)
DT  01-AUG-1988  (Rel. 08, Last sequence update)
DT  16-OCT-2001  (Rel. 40, Last annotation update)
DE  Sericin precursor (Silk gum protein).
DE  Bombyx mori (Silk moth).
OC  Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC  Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC  Ditrysia; Bombycoidea; Bombycidae; Bombyx.
OX  NCBI_TaxID=7091;
[1]
RN  SEQUENCE FROM N.A.
RP  TISSUE=Middle silk gland;
RC  MEDLINE=83082839; PubMed=6294094;
RX  Okamoto H., Ishikawa E., Suzuki Y.;
RA  "Structural analysis of sericin genes. Homologies with fibroin gene
RT  in the 5' flanking nucleotide sequences.";
RL  J. Biol. Chem. 257:15192-15199(1982).
RN  [2]
RP  SEQUENCE OF 317-354 FROM N.A.
RX  MEDLINE=87076763; PubMed=3024742;
RA  Michaille J.J., Coublie P., Prudhomme J.-C., Garel A.;
RT  "A single gene produces multiple sericin messenger RNAs in the silk
RL  gland of Bombyx mori.";
RL  Biochimie 68:1165-1173(1986).
CC  -!- FUNCTION: PROVIDES THE SILK FIBROIN THREAD WITH A STICKY COATING.
CC  -!- ACTS AS A CEMENT BY STICKING SILK THREADS TOGETHER.
CC  -!- SUBCELLULAR LOCATION: Secreted.
CC  -!- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE MIDDLE (MSG)
CC  SECTION OF SILK GLANDS.

```

```

*DR  ENBL; J01033; -; NOT_ANNOTATED_CDS.
DR   ENBL; J01034; -; NOT_ANNOTATED_CDS.
DR   ENBL; M26101; AAA27843.1; -;
DR   ENBL; M26102; AAA27844.1; -;
KW   Slik; Signal.
FT   SIGNAL 1 15 POTENTIAL.
FT   CHAIN 16 389 SERICIN.
SQ   SEQUENCE 389 AA; 39820 MW; BA605BC0305EAF19 CRC64;

Query Match 5.8%; Score 95.5; DB 1; Length 389;
Best Local Similarity 20.1%; Pred. No. 1.8;
Matches 65; Conservative 37; Mismatches 114; Indels 107; Gaps 11;

QY 9 STDONS-----VGNPDLNLRTRLVNSSGEIRTELTUKSRGPRGSKTGQOKOKKPTLRG 63
    |||::: ||| |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 136 STEESSSSSRAAASSTDASSNTDONSAGSSTSGGRTTYGYSNSRDG-----SVSS 188
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::

QY 64 MGAKLEFQRIEEKKOLAATVCDTSSVASISNNATRLPVPDYPGVVLQGFPSLLGSNR 123
    | | | | | : |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 189 TG-----SSSNTDSSNAGSSTSGGSTTYGYSNSR-----DGSYSTTGSSSNTDSNS 237
    | | | | | : |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::

QY 124 IYCGVGSGQWMDPVIPIWPGFVETSTTHLSSIS--NPQMFENASSNNRCDTCFKKR 180
    | | | | | |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 238 NSVGSRRG-----GSSSHEDSKSDENVTGSGSSNT----- 271
    | | | | | |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::

QY 181 LDGQNNVVRNSGGFSKYTWIPPPMNGYQYLLQSDHHQRSQCFLYDHRIRAAASVSAS 240
    | | | | | |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 272 -DNSNSVSGSTSGG-----RRTYGYSSNRSDGVSSTGSS 306
    | | | | | |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::

QY 241 STTINPYENATNHTGPMEEFGSYMEG---NPRNG-----SGG 275
    | | | | | |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 307 SNT----DNSNSVGSSTSGGSTTYGYSNSRSDGVSSTGSSSNTDSSNSAGSSTSGG 361
    | | | | | |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::

QY 276 VKYEFPFGKYGRVSVVATTSS 298
    | | | | | |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 362 SSTYGYSSNSHDGVSSTGSSGN 384
    | | | | | |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::

```

EMBL: U12141; AAA99658.1; -;
EMBL: 271330; CA95925.1; -;
SGD: S0004999; YNL054W.
Hypothetical protein.
SEQUENCE 1165 AA; 128140 MW; 0414978AA3066CE7 CRC64;
Query Match 5.8%; Score 95.5; DB 1; Length 1165;
Best Local Similarity 17.6%; Pred. No. 7.9; Mismatches 128; Indels 125; Gaps 14;
Matches 66; Conservative 55;
8 MSTDSVGNPNDLRLNTRLVVNSGEIRTEFLKGRKPGSKTQCKQKRLRGMGVA 67
527 ISTNPNL--NSVMQNNNLM---SGHHELDLSIKQEPHQL--QQQPP---MDVQ 575
68 KLERQRIEERKQKLAATVGTSSVASISNNATRLPVDPGVVLQGFPSLSGSRNRYCG 127
576 SVDSYTSNDPNDSNVIAKSPDKRSSVLSLK-----VSPHLLSS---TSSNGNTISCP 624
128 SVGSGQVWIDPVISPWGFEVTSSTHLSISNPOMFNASSNRCOTCFKKRLDGDQNN 187
625 NVATNSQELP-----NNDISTKSLSLRHSANRNSNYGDKRPLRTTVSK 674
188 VYRS--NGGFSKYTWIPPPMNGYDYLQ-----SDHHRSQGLYDHR--IARAASV 237
675 IFDSPNGAPLRRYSGVDPDVHVN--LEDYEQPHNYPTMNSVKKDEFYNSRNKKFPHGLNF 733
216 -----SDHRSQGLYDHR--IARAASV 237
734 YGDNVVEENNGDSSNVRPOHTNLQHEFIPEDNESDENDIHSFYNNKNDLETKPLI 793
238 S-----ASTTINPVFNATN-HTGPM-----EFGSGWENGNPANGSG 275
794 SDYGEDVDYDRPNATFNSYGSASNTHELPLHGRMPSRNNNDYDFMVGNNNTGNNQ 853
276 VKYEYFFCKYGER 289
854 LNEYTPFLRMKRGOR 867
RESULT 7
101 YEAST
R101_YEAST STANDARD; PRT; 625 AA.
P33400;
01-FEB-1994 (Rel. 28, Created)
01-FEB-1995 (Rel. 31, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Regulatory protein RIM101 (protein RIM1).
RIM101 OR RIM1 OR YHL027W.
Saccharomyces cerevisiae (Baker's yeast).
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A.
STRAIN=SK1.
MEDLINE=93376505; PubMed=8367297;
Su S.Y., Mitchell A.P.;
"Molecular characterization of the yeast meiotic regulatory gene
RIM1.";
Nucleic Acids Res. 21:3789-3797(1993).
[2]
SEQUENCE FROM N.A.
STRAIN=S288C/AB972;
MEDLINE=94378003; PubMed=8091229;
Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
Latreille P., Louis E.J., Macri C., Wardis E., Meneses S., Mouser L.,
Nhan M., Rifkin L., Riles L., St Peter H., Trevisan E., Vaughan K.,
Vignati D., Wilcox L., Wohlman P., Waterston R., Willson R.,
Vaudin M.;
"Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
VIII.";

Science 265:2077-2082(1994).
-!- FUNCTION: POSITIVE REGULATOR OF MEIOSIS THROUGH THE ACTIVATION OF
THE EXPRESSION OF IMEL. PROBABLE TRANSCRIPTION FACTOR.
-!- SUBCELLULAR LOCATION: Nuclear.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL: X72960; CA951462.1; -;
EMBL: U11583; AAB65039.1; -;
PIR: S38995; S38995.
PIR: S48941; S48941.
TRANSFAC: T01292; -;
SGD: S0001019; RIM101.
InterPro: IPR000822; Znf_C2H2.
Pfam: PF00096; zf-C2H2; 3.
PRINTS: PR00048; ZINCFINGER.
SMART: SM00355; Znf_C2H2; 3.
PROSITE: PS00028; ZINC_FINGER_C2H2_1; 2.
PROSITE: PS00157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
Metal-binding; Nuclear protein; Repeat; Meiosis.
FT ZN_FING 146 171
C2H2-TYPE.
FT ZN_FING 182 206
C2H2-TYPE.
FT ZN_FING 212 234
C2H2-TYPE.
FT DOMAIN 579 601
ASP/GLU-RICH (ACIDIC).
FT CONFLICT 68 68
N -> K (IN REF. 1).
FT CONFLICT 90 90
T -> S (IN REF. 1).
FT CONFLICT 111 111
S -> P (IN REF. 1).
FT CONFLICT 249 249
W -> L (IN REF. 1).
FT CONFLICT 261 261
C -> S (IN REF. 1).
FT CONFLICT 276 276
A -> G (IN REF. 1).
FT CONFLICT 308 308
Q -> QQQ (IN REF. 1).
FT CONFLICT 358 358
R -> Q (IN REF. 1).
FT CONFLICT 486 486
V -> I (IN REF. 1).
FT CONFLICT 519 519
G -> E (IN REF. 1).
SQ SEQUENCE 625 AA; 68232 MW; AD71E7B2D679E17E CRC64;
Query Match 5.8%; Score 95; DB 1; Length 625;
Best Local Similarity 21.4%; Pred. No. 3.8;
Matches 72; Conservative 39; Mismatches 139; Indels 86; Gaps 15;
QY 6 FFMSTDSVGNPNDLRLNTRLVVNSGEIRTEFLKGRKPGSKTQCKQKRLRGMG 65
Db 212 FGCSTCSKKFRPQDLKKHLKIHLESGILK---RKGKPGWSKRTSKKNK----- 259
QY 66 VAKLERQRIEERKQKLAATVGTSSVASISNNATRLPVDPGVVLQGFPSLSGSRNRY 125
Db 260 --SCASDAVSSCSASVPSAIGSFKSHSTSPQILPLPVGIS-----OHLPSQQQORALS 312
QY 126 CGGVSGGV-MIDPVISPMGFVETSTHLSISNPOMFNASS-----NNRC-----D 173
Db 313 LNLQCSDELSQYKPVYSP-----QLSARLQTLPLLYNNGSTVSQGSNSRMNVED 365
QY 174 TCFKKRLDGDQ-----NNVRSNGGFSKYT-----MIPPMNGYDYLQSDHH 219
Db 366 GCSNKTIANATQFTKLSRNMNNYILQSGSGSTESSSSSGRIPVAQTSYVQ-PPNAPSY 424
QY 220 QRSQGLYDHRIRAAASVSASSTTN-----PYFNEATNHTGPMEEFGSGWENGNR 270
Db 425 QSVQ-----GSSISATANTATYVVRVLAQYTPGSLTEHLPLPLHS----- 465
QY 271 NGSQGV--KEYEFPCKYGERVSVVATTSILYDCS 304
Db 466 NTAGGVNFQSQYAMPHY---PSVRAAPSYSSGCS 498
RESULT 8


```

13 NSVGNPNDLR-----NTRLVANSSEGEIETETLKSGRPGSKTGQOKKPK 59
1195 NNNNNNNLAMPFVFCOLREOGPIINAQMLAQROREILLSHQLRQLOLQOQOQOQOQOQOQ 1254
60 TLRGMGVAKLERQRIEKKOLAAATVDTSSVASISNNATRLPVDPGVVLQGFPSL 119
1255 -----QQOQOQOQOQOQOQOQOQHRRAMWRSOGLAMPNNVGS-----GIPASI 1300
120 GSNRIYCGV-----GSGQVMIDPVISPMGFVETSTTHELSSISNPFMFNASSNN 170
1301 NSPRIPOGSGTQFPFPNNVGTGIPSPFPPTSPFSVPVPSGQSLSHSS----- 1349
171 RCDTCFKKRLDQNN-----VVRNSGGGSKTYMIDPPMNGVDYQLLQSDHHQRSGF 225
1350 -----LHGSOMTLANQIGMSGGQGP-VYNNQMOHNAFQFANSQMSQSDPGF 1398
226 -----LYDHRIARAASVSASTTINPFNEATNHTGPMEEFGYMEGNPNRGS 273
1399 TGATTPQSIFMSFRCHIQSSMMQSQANPAY-----QSELNGRAQGNPAGNS 1446

RESULT 10
EWS_MOUSE STANDARD; PRT; 655 AA.
Q61545:
01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
RNA-binding protein EWS.
EWS OR EWSH.
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
TISSUE=Testis;
MEDLINE=95130099; Pubmed=7829090;
Plougastel B., Mattei M.-G., Thomas G., Delattre O.;
"Cloning and chromosome localization of the mouse Ews gene.";
Genomics 23:278-281(1994).
-!- FUNCTION: MIGHT FUNCTION AS A REPRESSOR (BY SIMILARITY).
-!- SUBCELLULAR LOCATION: Nuclear.
-!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
-!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
-!- SIMILARITY: CONTAINS 1 RANBP2-TYPE ZINC FINGER.
-!- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).

EMBL; X79233; CAA55815.1; .
MGD; MGI:99960; Ewsh.
InterPro: IPR000504; RNA_rec_mot.
InterPro: IPR001876; Znf_RanGDP.
Pfam: PF00076; rtm; 1.
Pfam: PF00641; zf-RanBP; 1.
SMART: SM00360; RRM; 1.
SMART: SM00547; Znf_RBZ; 1.
PROSITE: PS50096; IQ; FALSE_NEG.
PROSITE: PS50102; RRM; 1.
PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
PROSITE: PS01358; ZF_RANBP2_1; 1.
PROSITE: PS01999; ZF_RANBP2_2; 1.
Nuclear protein; zinc-finger; Zinc; RNA-binding; Metal-binding;
Calmodulin-binding; Repeat; Methylation; Phosphorylation;
Transcription regulation; Repressor.
DOMAIN 1 285 EAD (GLN/PRO/THR-RICH).

```

Query Match	5.7%	Score 93;	DB 1;	Length 655;
Best Local Similarity	21.8%	Pred. No. 5.8;		

DR InterPro: IPR000201; DNaPol viral N.

Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brokstein P., Bottlier P.,
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Duan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Maitel B., McIntosh T.C., McCleod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
"The genome sequence of Drosophila melanogaster";
Science 287:2185-2195(2000).
-!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE REGULATION OF
MICROTUBULE ASSEMBLY AND INTERACTION.
-!- SUBCELLULAR LOCATION: ASSOCIATED WITH CYTOPLASMIC MICROTUBULES AND
WITH THE MITOTIC SPINDLE.
-!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; C2, B3 AND J5 (SHOWN HERE); ARE
PRODUCED BY ALTERNATIVE SPLICING.
-!- MISCELLANEOUS: PHOSPHORYLATION OF VARIOUS SERINE RESIDUES MAY PLAY
A REGULATORY ROLE. THE BASIC DOMAIN CONTAINS NUMEROUS SEQUENCES
THAT MATCH KNOWN CONSENSUS SEQUENCES OF SEVERAL DIFFERENT PROTEIN
KINASES.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X54061; CAA37996.1; -;
EMBL; AE003780; AAF57214.1; -;
PIR; A36685; A36685.
R FlyBase: FBgn002645; Map205.
T Microtubules; Alternative splicing: Phosphorylation.
T DOMAIN 1 784 ASP/GLU-RICH (ACIDIC).
T DOMAIN 785 1124 ARG/LYS-RICH (BASIC).
T DOMAIN 1125 1185 ASP/GLU-RICH (ACIDIC).
T BINDING 745 977 TO MICROTUBULES (POTENTIAL).
T VARSPLIC 557 578 MISSING (IN ISOFORM B3 AND ISOFORM C2).
T VARSPLIC 650 703 MISSING (IN ISOFORM C2).
T VARSPLIC 704 704 D -> N (IN ISOFORM C2).
Q SEQUENCE 1185 AA; 126669 MW; 47B422E2CE03F70 CRC64;
Query Match 5.5%; Score 90.5; DB 1; Length 1185;
Best Local Similarity 20.8%; Pred. No. 21;
Matches 65; Conservative 37; Mismatches 146; Indels 65; Gaps 13;
9 STDQNSVGNPNLLRLNRLVWSSGGIETRTLSRGRKP-----GSKTGQKQKK 58

Db 862 STTSLLGNPKRSLSS-----NVGSTVAPPT-KLSGTRPATAPVSKVTLGAKT-----ITNK 912
Qy 59 PTLRGMGVAKLQRIEERIEEKQLAAATVGDTSVASISNNATRLPVPVDPGVVLQGFSS 118
Db 913 PTASGTASDNVTRITRLPLVSTNARRPA--ISGTGVSASSTARRPVTNAKG-----SAFGS 966
Qy 119 LGSNRIYCGGVSGQVMIDPVLSPWGFVETSTTHLSSISNP-----QMFNASSN 169
Db 967 AASTKVRPAATMTAPVK-PKVLSPRSTISSTTVTKVSTPTSPFSFSTRPNKQKQSLGK 1025
Qy 170 NRCDFCFKKRLDGDQNNVRS-----NGGFSKYTMIPPPMGVYDVLLOSD 217
Db 1026 NTSSTTTATATITKSTAKSPAKFTSHASLTYNNGTSRRLVLP-----GSSSTTTTSS 1081
Qy 218 HHORGGLYDHRIARAASVSASSTTINPYNEATNHTGP--MEEFSGSYMCGNPRNGSGG 275
Db 1082 LRKSP-----LKAAPGKAASKPLPQSKDGTAKSPAVLKARNSTLMG-----GESV 1129
Qy 276 VKYEYFFPKYGE 288
Db 1130 APSNECVPTNGQ 1142
RESULT 15
NCO2_HUMAN
ID NCO2_HUMAN STANDARD; PRT; 1464 AA.
AC Q15596;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear receptor coactivator 2 (transcriptional intermediary factor
2).
GN NCOA2 OR TIF2
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=96312964; PubMed=86708070;
RA Voegel J.J., Heine M.J.S., Zechel C., Chambon P., Gronemeyer H.;
RT "TIF2, a 160 kDa transcriptional mediator for the ligand-dependent
activation function AF-2 of nuclear receptors.";
RL EMBO J. 15:3657-3675(1996).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=98094372; PubMed=9430642;
RA Voegel J.J., Heine M.J.S., Tini M., Vivat V., Chambon P.,
Gronemeyer H.;
RT "The coactivator TIF2 contains three nuclear receptor-binding motifs
and mediates transactivation through CBP binding-dependent and
-independent pathways.";
RL EMBO J. 17:507-519(1998).
CC -!- FUNCTION: TRANSCRIPTIONAL COACTIVATOR FOR STEROID RECEPTORS AND
NUCLEAR RECEPTORS. COACTIVATOR OF THE STEROID BINDING DOMAIN
(AF-2) BUT NOT OF THE MODULATING N-TERMINAL DOMAIN (AF-1).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; X97674; CAA66263.1; -;
DR TRANSFAC; T02483; -;
DR Genew; HGNC:7669; NCOA2.
DR MIM; 601993; -;

```

DR InterPro; IPR001092; HLH_basic.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF00989; PAS; 1.
DR SMART; SM00353; HLH; 1.
DR SMART; SM00086; PAC; 1.
DR SMART; SM00091; PAS; 1.
DR PROSITE; PS00888; HLH_2; 1.
DR PROSITE; PS0112; PAS; 1.
KW Transcription regulation; Activator; Nuclear protein.
FT DOMAIN 119 183
FT DOMAIN 1254 1260
SQ SEQUENCE 1464 AA; 159156 MW; 0A61AA5D1878304B CRC64;

Query Match      5.5%; Score 90.5; DB 1; Length 1464;
Best Local Similarity 19.9%; Pred. NO. 27;
Matches 52; Conservative 36; Mismatches 88; Indels 85; Gaps 9;

QY 13 NSVGNPNLLR-----NRLVNVSSGEIRTELKSRGKPGSKTGOQKQKPTLRG 63
Db 1211 SNVSNVLTLPQVPTQAPINAOQLAQOREILNQHLRQRM---HQOQVQOQTLMRG 1267

QY 64 MGVAKLERQRIEKKOLAAATVGDTSVASISN-----NATRLPVPVDPGVVLOGFPS 117
Db 1268 QGL-----NMTPSWAPSGMPATMSNPRIPOANAQQPPFPNYGISQOQDPG 1314

QY 118 SLGSNRIYCGVGSGQWIDPVISPMGFVETSTTHELSSISNPQMFNASSNNRCDTCK 177
Db 1315 FTGAT-----TPQSPLMSP-----RMAHTQSPMMQOQOAN-----P 1345

QY 178 KKRLDQDNVVRNSGGGFSKYTMIPPPMNGYDQYLLOSDHHQRSGGFLYDRIARAASV 237
Db 1346 AYQAPSDINGWAQGNMGNSMFSQSQSPPHFG-----QOANT 1381

QY 238 SASSTINPYFNEATNHTGPM 258
Db 1382 SMYSNNMNINVSMTN-TGGM 1401

```

Search completed: January 26, 2003, 00:04:15
Job time : 34 secs